EXPRESS MAIL MAILING LABEL NO. EL653442545 US

PATENT
Attorney Docket No.: RIB-001CP

RIBOSOME STRUCTURE AND PROTEIN SYNTHESIS INHIBITORS

RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application No. 09/635,708, filed August 9, 2000, and claims the benefit of (i) U.S. Provisional Application No. 60/223,977, filed August 9, 2000, (ii) U.S. Provisional Application No. [Atty. Docket No. RIB-002PR], entitled "The Kink-Turn: a New RNA Secondary Structure Motif," filed July 20, 2001, and (iii) U.S. Provisional Application No. [Atty. Docket No. RIB-002PR2], entitled "The Kink-Turn: a New RNA," filed August 1, 2001, the disclosures of each of the foregoing of which are incorporated by reference herein.

GOVERNMENT LICENSE RIGHTS

Certain work described herein was supported, in part, by Federal Grant Nos. NIH-GM22778 and NIH-GM54216, awarded by the National Institutes of Health. The Government may have certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates generally to the field of protein biosynthesis and to modulators, for example, inhibitors, of protein biosynthesis. More particularly, the invention relates to methods and compositions for elucidating the three-dimensional structure of the large ribosomal subunit, either alone or in combination with a protein synthesis inhibitor; the three-dimensional structure of the large ribosomal subunit, either alone or in combination with a protein synthesis inhibitor; the use of such structures in the design and testing of novel protein synthesis inhibitors; and novel protein synthesis inhibitors.

BACKGROUND

I. Ribosomes: Structure, Function, and Composition

Ribosomes are ribonucleoproteins which are present in both prokaryotes and eukaryotes. They comprise about two-thirds RNA and one-third protein. Ribosomes are the cellular organelles responsible for protein synthesis. During gene expression, ribosomes translate the genetic information encoded in a messenger RNA into protein (Garrett *et al.* (2000) "*The Ribosome: Structure, Function, Antibiotics and Cellular Interactions,*" American Society for Microbiology, Washington, D.C.).

Ribosomes comprise two nonequivalent ribonucleoprotein subunits. The larger subunit (also known as the "large ribosomal subunit") is about twice the size of the smaller subunit (also known as the "small ribosomal subunit"). The small ribosomal subunit binds messenger RNA (mRNA) and mediates the interactions between mRNA and transfer RNA (tRNA) anticodons on which the fidelity of translation depends. The large ribosomal subunit catalyzes peptide bond formation -- the peptidyl-transferase reaction of protein synthesis -- and includes (at least) two different tRNA binding sites: the A-site which accommodates the incoming aminoacyl-tRNA, which is to contribute its amino acid to the growing peptide chain, and the P-site which accommodates the peptidyl-tRNA complex, *i.e.,* the tRNA linked to all the amino acids that have so far been added to the peptide chain. The large ribosomal subunit also includes one or more binding sites for G-protein factors that assist in the initiation, elongation, and termination phases of protein synthesis. The large and small ribosomal subunits behave independently during the initiation phase of protein synthesis; however, they assemble into complete ribosomes when elongation is about to begin.

The molecular weight of the prokaryotic ribosome is about 2.6×10^6 daltons. In prokaryotes, the small ribosomal subunit contains a 16S (Svedberg units) ribosomal RNA (rRNA) having a molecular weight of about 5.0×10^5 daltons. The large ribosomal subunit contains a 23S rRNA having a molecular weight of about 1.0×10^6 daltons and a 5S rRNA having a molecular weight of about 4.0×10^5 daltons. The prokaryotic small subunit contains about 20 different proteins and its large subunit contains about 35 proteins. The large and small ribosomal subunits together constitute a 70S ribosome in prokaryotes.

Eukaryotic ribosomes generally are bigger than their prokaryotic counterparts. In eukaryotes, the large and small subunits together make an 80S ribosome. The small subunit of a eukaryotic ribosome includes a single 18S rRNA, while the large subunit includes a 5S rRNA, a 5.8S rRNA, and a 28S rRNA. The 5.8S rRNA is structurally related to the 5' end of the prokaryotic 23S rRNA, and the 28S rRNA is structurally related to the remainder of the prokaryotic 23S rRNA (Moore (1998) *Annu. Rev. Biophys.* 27: 35-58). Eukaryotic ribosomal proteins are qualitatively similar to the prokaryotic ribosomal proteins; however, the eukaryotic proteins are bigger and more numerous (Moore (1998) *supra*).

II. Structural Conservation of the Large Ribosomal Subunit

While the chemical composition of large ribosomal subunits vary from species to species, the sequences of their components provide unambiguous evidence that they are similar in three-dimensional structure, function in a similar manner, and are related evolutionarily. The evolutionary implications of rRNA sequence data available are reviewed in the articles of Woese and others in part II of Ribosomal RNA. Structure, Evolution, Processing and Function in Protein Biosynthesis, (Zimmermann and Dahlberg, eds., CRC Press, Boca Raton, FL, 1996). The article by Garret and Rodriguez-Fonseca in part IV of the same volume discusses the unusually high level of sequence conservation observed in the peptidyl transferase region of the large ribosomal subunit. The ribosomes of archeal species like *Haloarcula marismortui* resemble those obtained from eubacterial species like *E. coli* in size and complexity. However, the proteins in *H. marismortui* ribosomes are more closely related to the ribosomal proteins found in eukaryotes (Wool *et al.* (1995) *Biochem. Cell Biol.* 73: 933-947).

III. Determination of the Structure of Ribosomes

Much of what is known about ribosome structure is derived from physical and chemical methods that produce relatively low-resolution information. Electron microscopy (EM) has contributed to an understanding of ribosome structure ever since the ribosome was discovered. In the 1970s, low resolution EM revealed the shape and quaternary organization of the ribosome. By the end of 1980s, the positions of the surface epitopes of all the proteins in the *E. coli* small subunit, as well as many in the large subunit, had been mapped using immunoelectron microscopy techniques (Oakes *et al.* (1986), Structure, Function and Genetics of Ribosomes, (Hardesty, B. and Kramer, G., eds.) Springer-Verlag, New York, NY, pp. 47-67; Stoeffler *et al.*

(1986), Structure, Function and Genetics of Ribosomes, (Hardesty, B. and Kramer, G., eds.) Springer-Verlag, New York, NY, pp.28-46). In the last few years, advances in single-particle cryo-EM and image reconstruction have led to three-dimensional reconstructions of the *E. coli* 70S ribosome and its complexes with tRNAs and elongation factors to resolutions of between 15 Å and 25 Å (Stark *et al.* (1995) *Structure* 3: 815-821; Stark *et al.* (1997) *Nature* 3898: 403-406; Agrawal *et al.* (1996) *Science* 271: 1000-1002; Stark *et al.* (1997) *Cell* 28: 19-28). Additionally, three-dimensional EM images of the ribosome have been produced at resolutions sufficiently high so that many of the proteins and nucleic acids that assist in protein synthesis can be visualized bound to the ribosome. An approximate model of the RNA structure in the large subunit has been constructed to fit a 7.5 Å resolution electron microscopic map of the 50S subunit from *E. coli* and available biochemical data (Mueller *et al.* (2000) *J. Mol. Biol.* 298: 35-59).

While the insights provided by EM have been useful, it has long been recognized that a full understanding of ribosome structure would derive only from X-ray crystallography. In 1979, Yonath and Wittman obtained the first potentially useful crystals of ribosomes and ribosomal subunits (Yonath *et al.* (1980) *Biochem. Internat.* 1: 428-435). By the mid 1980s, scientists were preparing ribosome crystals for X-ray crystallography (Maskowski *et al.* (1987) *J. Mol. Biol.* 193: 818-822). The first crystals of 50S ribosomal subunit from *H. marismortui* were obtained in 1987. In 1991, improvements were reported in the resolution of the diffraction data obtainable from the crystals of the 50S ribosomal subunit of *H. marismortui* (van Bohlen, K. (1991) *J. Mol. Biol.* 222: 11).

In 1995, low resolution electron density maps for the large and small ribosomal subunits from halophilic and thermophilic sources were reported (Schlunzen *et al.* (1995) *Biochem. Cell Biol.* 73: 739-749). However, these low resolution electron density maps proved to be spurious (Ban *et al.* (1998) *Cell* 93: 1105-1115).

The first electron density map of the ribosome that showed features recognizable as duplex RNA was a 9 Å resolution X-ray crystallographic map of the large subunit from *Haloarcula marismortui* (Ban *et al.* (1998) *supra*). Extension of the phasing of that map to 5 Å resolution made it possible to locate several proteins and nucleic acid sequences, the structures of which had been determined independently (Ban *et al.* (1999) *Nature* 400: 841-847).

At about the same time, using similar crystallographic strategies, a 7.8 Å resolution map was generated of the entire *Thermus thermophilus* ribosome showing the positions of tRNA molecules bound to its A-, P-, and E- (protein exit site) sites (Cate *et al.* (1999) *Science* 285: 2095-2104), and a 5.5 Å resolution map of the 30S subunit from *T. thermophilus* was obtained that allowed the fitting of solved protein structures and the interpretation of some of its RNA features (Clemons, Jr. *et al.* (1999) *Nature* 400: 833-840). Subsequently, a 4.5 Å resolution map of the *T. thermophilus* 30S subunit was published, which was based in part on phases calculated from a model corresponding to 28% of the subunit mass that had been obtained using a 6 Å resolution experimental map (Tocilj *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96: 14252-14257).

IV. Location of the Peptidyl Transferase Site in the Large Ribosomal Subunit

It has been known for about 35 years that the peptidyl transferase activity responsible for the peptide bond formation that occurs during messenger RNA-directed protein synthesis is intrinsic to the large ribosomal subunit (Traut *et al.* (1964) *J. Mol. Biol.* 10: 63; Rychlik (1966) *Biochim. Biophys. Acta* 114: 425; Monro (1967) *J. Mol. Biol.* 26: 147-151; Maden *et al.* (1968) *J. Mol. Biol.* 35: 333-345) and it has been understood for even longer that the ribosome contains proteins as well as RNA. In certain species of bacteria, for example, the large ribosomal subunit contains about 35 different proteins and two RNAs (Noller (1984) *Ann. Rev. Biochem.* 53: 119-162; Wittmann-Liebold *et al.* (1990) The Ribosome: Structure, Function, and Evolution, (W.E. Hill et al., eds.) American Society for Microbiology, Washington, D.C. (1990), pp. 598-616). These findings posed three related questions. Which of the almost 40 macromolecular components of the large ribosomal subunit contribute to its peptidyl transferase site, where is that site located in the large subunit, and how does it work?

By 1980, the list of components that might be part of the ribosome's peptidyl transferase had been reduced to about half a dozen proteins and 23S rRNA (see Cooperman (1980) Ribosomes: Structure, Function and Genetics, (G. Chambliss et al., eds.) University Park Press, Baltimore, MD (1980), 531-554), and following the discovery of catalytic RNAs (Guerrier-Takada *et al.* (1983) *Cell* 35: 849-857; Kruger *et al.* (1982) *Cell* 31: 147-157), the hypothesis that 23S rRNA might be its sole constituent, which had been proposed years earlier, began to gain favor. In 1984, Noller and colleagues published affinity labeling results which showed that U2619 and U2620 (in *E. coli*: U2584, U2585) are adjacent to the CCA-end of P-site-bound

tRNA (Barta *et al.* (1984) *Proc. Nat. Acad. Sci. USA* 81: 3607-3611; Vester *et al.* (1988) *EMBO J.* 7: 3577-3587). These nucleotides appear to be part of a highly conserved internal loop in the center of domain V of 23S rRNA. The hypothesis that this loop is intimately involved in the peptidyl transferase activity was supported by the observation that mutations in that loop render cells resistant to many inhibitors of peptidyl transferase, and evidence implicating it in this activity has continued to mount (see, Noller (1991) *Ann. Rev. Biochem.* 60: 191-227; Garrett *et al.* (1996) Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis, (R.A. Zimmerman and A.E. Dahlberg, eds.) CRC Press, Boca Raton, FL (1996), pp. 327-355).

Definitive proof that the central loop in domain V is the sole component of the ribosome involved in the peptidyl transferase activity has remained elusive, however. Studies have shown that it was possible to prepare particles that retained peptidyl transferase activity by increasingly vigorous deproteinizations of large ribosomal subunits, however, it was not possible to produce active particles that were completely protein-free. Nevertheless, combined with earlier reconstitution results (Franceschi *et al.* (1990) *J. Biol. Chem.* 265: 6676-6682), this work reduced the number of proteins that might be involved to just two: L2 and L3 (see, Green *et al.* (1997) *Annu. Rev. Biochem.* 66: 679-716). More recently, Watanabe and coworkers reported success in eliciting peptidyl transferase activity from *in vitro* synthesized, protein-free 23S rRNA (Nitta *et al.* (1998) *RNA* 4: 257-267), however, their observations appear not to have withstood further scrutiny. Thus the question still remained: is the ribosome a ribozyme or is it not?

Over the years, the location of the peptidyl transferase site in the ribosome has been approached almost exclusively by electron microscopy. In the mid-1980s evidence that there is a tunnel running through the large ribosomal subunit from the middle of its subunit interface side to its back (Milligan *et al.* (1986) *Nature* 319: 693-695; Yonath *et al.* (1987) *Science* 236: 813-816) began to accumulate, and there has been strong reason to believe that polypeptides pass through it as they are synthesized (Bernabeu *et al.* (1982) *Proc. Nat. Acad. Sci. USA* 79: 3111-3115; Ryabova *et al.* (1988) *FEBS Letters* 226: 255-260; Beckmann *et al.* (1997) *Science* 278: 2123-2126). More recent cryo-EM investigations (Frank *et al.* (1995) *Nature* 376: 441-444; Frank *et al.* (1995) *Biochem. Cell Biol.* 73: 757-765; Stark *et al.* (1995) *supra*) confirmed the existence of the tunnel and demonstrated that the CCA-ends of ribosome-bound tRNAs bound to

the A- and P-sites are found in the subunit interface end of the tunnel. Consequently, the peptidyl transferase site must be located at that same position, which is at the bottom of a deep cleft in the center of the subunit interface surface of the large subunit, immediately below its central protuberance.

The substrates of the reaction catalyzed at the peptidyl transferase site of the large subunit are an aminoacyl-tRNA (aa-tRNA) and a peptidyl-tRNA. The former binds in the ribosome's A-site and the latter in its P-site. The α-amino group of the aa-tRNA attacks the carbon of the carbonyl acylating the 3' hydroxyl group of the peptidyl-tRNA, and a tetrahedral intermediate is formed at the carbonyl carbon. The tetrahedral intermediate resolves to yield a peptide extended by one amino acid esterified to the A-site bound tRNA and a deacylated tRNA in the P-site.

This reaction scheme is supported by the observations of Yarus and colleagues who synthesized an analogue of the tetrahedral intermediate by joining an oligonucleotide having the sequence CCdA to puromycin via a phosphoramide group (Welch *et al.* (1995) *Biochemistry* 34: 385-390). The sequence CCA, which is the 3' terminal sequence of all tRNAs, binds to the large subunit by itself, consistent with the biochemical data showing that the interactions between tRNAs and the large subunit largely depend on their CCA sequences (Moazed *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88: 3725-3728). Puromycin is an aa-tRNA analogue that interacts with the ribosomal A-site, and the phosphoramide group of the compound mimics the tetrahedral carbon intermediate. This transition state analogue, CCdA-phosphate-puromycin (CCdA-p-Puro), binds tightly to the ribosome, and inhibits its peptidyl transferase activity (Welch *et al.* (1995) *supra*).

V. Structure Determination of Macromolecules Using X-ray Crystallography

In order to better describe efforts undertaken to determine the structure of ribosomes, a general overview of X-ray crystallography is provided below.

Each atom in a crystal scatters X-rays in all directions, but crystalline diffraction is observed only when a crystal is oriented relative to the X-ray beam so that the atomic scattering interferes constructively. The orientations that lead to diffraction may be computed if the wavelength of the X-rays used and the symmetry and dimensions of the crystal's unit cell are known (Blundell *et al.* (1976) Protein Crystallography (Molecular Biology Series), Academic Press, London). The result is that if a detector is placed behind a crystal that is being irradiated

with monochromatic X-rays of an appropriate wavelength, the diffraction pattern recorded will consist of spots, each spot representing one of the orientations that gives rise to constructive interference.

Each spot in such a pattern, however it is recorded, is characterized by (i) an intensity (its blackness); (ii) a location, which encodes the information about diffraction orientation; and (iii) a phase. If all of those things are known about each spot in a crystal diffraction pattern, the distribution of electrons in the unit cell of the crystal may be computed by Fourier transformation (Blundell *et al.* (1976) *supra*), and from that distribution or electron density map, atomic positions can be determined.

Unfortunately, the phase information essential for computing electron distributions cannot be measured directly from diffraction patterns. One of the methods routinely used to determine the phases of macromolecules, such as proteins and nucleic acids, is called multiple isomorphous replacement (MIR) which involves the introduction of new X-ray scatterers into the unit cell of the crystal. Typically, these additions are heavy atoms, which make a significant contribution to the diffraction pattern. It is important that the additions be sufficiently low in number so that their positions can be located and that they leave the structure of the molecule or of the crystal cell unaltered, *i.e.,* the crystals should be isomorphous. Isomorphous replacement usually is performed by diffusing different heavy-metal complexes into the channels of the preformed protein crystals. Macromolecules expose side chains (such as SH groups) in these solvent channels that are able to bind heavy metals. It is also possible to replace endogenous light metals in metalloproteins with heavier ones, *e.g.,* zinc by mercury, or calcium by samarium. Alternatively, the isomorphous derivative can be obtained by covalently attaching a heavy metal to the macromolecule in solution and then subjecting it to crystallization conditions.

Heavy metal atoms routinely used for isomorphous replacement include but are not limited to mercury, uranium, platinum, gold, lead, and selenium. Specific examples include mercury chloride, ethyl-mercury phosphate, and osmium pentamine, iridium pentamine. Since such heavy metals contain many more electrons than the light atoms (H, N, C, O, and S) of the protein, the heavy metals scatter x-rays more strongly. All diffracted beams would therefore increase in intensity after heavy-metal substitution if all interference were positive. In fact,

however, some interference is negative; consequently, following heavy-metal substitution, some spots increase in intensity, others decrease, and many show no detectable difference.

Phase differences between diffracted spots can be determined from intensity changes following heavy-metal substitution. First, the intensity differences are used to deduce the positions of the heavy atoms in the crystal unit cell. Fourier summations of these intensity differences give maps, of the vectors between the heavy atoms, the so-called Patterson maps. From these vector maps, the atomic arrangement of the heavy atoms is deduced. From the positions of the heavy metals in the unit cell, the amplitudes and phases of their contribution to the diffracted beams of protein crystals containing heavy metals is calculated.

This knowledge then is used to find the phase of the contribution from the protein in the absence of the heavy-metal atoms. As both the phase and amplitude of the heavy metals and the amplitude of the protein alone is known, as well as the amplitude of the protein plus heavy metals (*i.e.,* protein heavy-metal complex), one phase and three amplitudes are known. From this, the interference of the X-rays scattered by the heavy metals and protein can be calculated to determine if the interference is constructive or destructive. The extent of positive or negative interference, with knowledge of the phase of the heavy metal, give an estimate of the phase of the protein. Because two different phase angles are determined and are equally good solutions, a second heavy-metal complex can be used which also gives two possible phase angles. Only one of these will have the same value as one of the two previous phase angles; it therefore represents the correct phase angle. In practice, more than two different heavy-metal complexes are usually made in order to give a reasonably good estimate of the phase for all reflections. Each individual phase estimate contains experimental errors arising from errors in the measured amplitudes. Furthermore, for many reflections, the intensity differences are too small to measure after one particular isomorphous replacement, and others can be tried.

The amplitudes and the phases of the diffraction data from the protein crystals are used to calculate an electron-density map of the repeating unit of the crystal. This map then is interpreted to accommodate the residues of the molecule of interest. That interpretation is made more complex by several limitations in the data. First, the map itself contains errors, mainly due to errors in the phase angles. In addition, the quality of the map depends on the resolution of the diffraction data, which, in turn, depends on how well-ordered the crystals are. This directly

influences the quality of the map that can be produced. The resolution is measured in angstrom units (Å); the smaller this number is, the higher the resolution and, therefore, the greater the amount of detail that can be seen.

Building the initial model is a trial-and-error process. First, one has to decide how a polypeptide chain or nucleic acid weaves its way through the electron-density map. The resulting chain trace constitutes a hypothesis by which one tries to match the density of side chains to the known sequence of the polypeptide or nucleic acid. When a reasonable chain trace has finally been obtained, an initial model is built that fits the atoms of the molecule into the electron density. Computer graphics are used both for chain tracing and for model building to present the data and manipulated the models.

The initial model will contain some errors. Provided the crystals diffract to high enough resolution (*e.g.,* better than 3.5 Å), most or substantially all of the errors can be removed by crystallographic refinement of the model using computer algorithms. In this process, the model is changed to minimize the difference between the experimentally observed diffraction amplitudes and those calculated for a hypothetical crystal containing the model (instead of the real molecule). This difference is expressed as an R factor (residual disagreement) which is 0.0 for exact agreement and about 0.59 for total disagreement.

In general, the R factor for a well-determined macromolecular structure preferably lies between 0.15 and 0.35 (such as less than about 0.24-0.28). The residual difference is a consequence of errors and imperfections in the data. These derive from various sources, including slight variations in the conformation of the protein molecules, as well as inaccurate corrections both for the presence of solvent and for differences in the orientation of the microcrystals from which the crystal is built. This means that the final model represents an average of molecules that are slightly different both in conformation and orientation.

In refined structures at high resolution, there are usually no major errors in the orientation of individual residues, and the estimated errors in atomic positions are usually around 0.1-0.2 Å, provided the sequence of the protein or nucleic acid is known. Hydrogen bonds, both within the molecule of interest and to bound ligands, can be identified with a high degree of confidence.

Typically, X-ray structures can be determined provided the resolution is better than 3.5 Å. Electron-density maps are interpreted by fitting the known amino acid and/or nucleic acid sequences into regions of electron density.

VI. The Need for Higher Resolution for the 50S Ribosomal Subunit

Although the art provides crystals of the 50S ribosomal subunit, and 9Å and 5 Å resolution X-ray crystallographic maps of the structure of the 50S ribosome, the prior art crystals and X-ray diffraction data are not sufficient to establish the three-dimensional structures of all 31 proteins and 3,043 nucleotides of the 50S ribosomal subunit. Thus, the prior art crystals and maps are inadequate for the structure-based design of active agents, such as herbicides, drugs, insecticides, and animal poisons.

More detailed, higher resolution X-ray crystallographic maps are necessary in order to determine the location and three-dimensional structure of the proteins and nucleotides in ribosomes and ribosomal subunits, particularly for the 50S ribosomal subunit. An accurate molecular structure of the 50S ribosomal subunit will not only enable further investigation and understanding of the mechanism of protein synthesis, but also the development of effective therapeutic agents and drugs that modulate (*e.g.,* induce or inhibit) protein synthesis.

SUMMARY OF THE INVENTION

The present invention is based, in part, upon the determination of a high resolution atomic structure of a ribosomal subunit, more particularly, a large subunit of a ribosome. The high resolution structure has been determined for a large ribosomal subunit present in the organism, *Haloarcula marismortui*. However, in view of the high level of sequence and structural homology between ribosomes of organisms in different kingdoms, the structural information disclosed herein can be used to produce, using routine techniques, high resolution structural models of large ribosomal units for any organism of interest.

Although there is significant homology between ribosomes of different organisms, for example, between ribosomes of humans and certain human pathogens, there still are differences that can be exploited therapeutically. For example, many clinically and commercially significant protein synthesis inhibitors, for example, antibiotics such as streptomycin, tetracycline, chloramphenicol and erythromycin, selectively target bacterial ribosomes and disrupt bacterial

protein synthesis but at the same time do not target or otherwise significantly affect human ribosome function. As a result, over the years antibiotics have proven to be invaluable in the treatment of microbial infections in humans. However, there is still an ongoing need for new protein synthesis inhibitors, particularly because of the development of strains of pathogens that are resistant to known antibiotics. The information provided herein provides insights into the design of new protein synthesis inhibitors.

The invention herein provides methods and compositions for resolving to a high resolution, the three-dimensional structure of a ribosomal subunit of interest. In addition, the invention provides computer systems containing atomic co-ordinates that define at least a portion of the three-dimensional structure of a ribosome, more specifically, a large ribosomal subunit. In addition, the invention provides methods of using the atomic co-ordinates to identify new molecules that can selectively bind ribosomes, and that preferably act as selective inhibitors of protein synthesis. In addition, the invention provides new families of protein synthesis inhibitors. Each of these aspects of the invention is discussed in more detail below.

In one aspect, the invention provides crystals, preferably untwinned crystals, of ribosomes or ribosomal subunits that have an average thickness greater than about 15 μm. More specifically, the invention provides crystals having an average thickness from about 16 μm to about 65 μm, or from about 66 μm to about 105 μm, or from about 104 μm to about 155 μm, or from about 156 μm to about 205 μm. In particular, the invention provides crystals having an average thickness from about 100 μm to about 200 μm.

In a preferred embodiment, the invention provides crystals that have an average thickness greater than about 15 μm and/or are untwinned wherein the crystals comprise the large ribosomal subunit. More particularly, the present invention provides such crystals wherein the large ribosomal subunit is a 50S or 60S ribosomal subunit. The crystals may be obtained using the ribosomes or ribosomal subunits from prokaryotes or from eukaryotes. In a preferred embodiment, the invention provides crystals containing ribosomes or ribosomal subunits obtained from bacteria or archaebacteria, more particularly, from the organism *Haloarcula marismortui*. However, crystals may be obtained from ribosomes or ribosomal subunits from any organism, particularly from animals, more particularly from mammals, and even more particularly from humans.

In another preferred embodiment, the crystals diffract X-rays to a resolution of at least about 4.5 Å, more preferably to a resolution of at least about 3.0 Å, and most preferably to a resolution of about 2.4 Å for the determination of atomic co-ordinates of ribosomes or ribosomal subunits. In another preferred embodiment, the crystals of the present invention may also include a ligand, for example, a protein synthesis inhibitor, for example, an antibiotic, (such as a macrolide antibiotic) complexed with, or bound to a ribosome or ribosomal subunit.

In another aspect, the invention provides crystals of 50S ribosomal subunits whose atomic structure is characterized by the atomic co-ordinates deposited at the Protein Data Bank ID: 1FFK or 1JJ2. The invention further provides phases computed from the co-ordinates of the deposited co-ordinates and the uses of such phase information. In a preferred embodiment, the invention provides crystals of 50S ribosomal subunits whose atomic structure is characterized by the atomic co-ordinates deposited at the Protein Data Bank ID: 1FFZ (large ribosomal subunit complexed with CCdA-p-Puro); or 1FG0 (large ribosomal subunit complexed with a mini-helix analogue of aminoacyl-tRNA); as well as those ribosomal subunits whose atomic structure is characterized by the atomic coordinates listed in a file contained on Disk No. 3 of 3, specifically: large ribosomal subunit complexed with anisomycin (file name: anisomysin.pdb); large ribosomal subunit complexed with blasticidin (file name: blasticidin.pdb); large ribosomal subunit complexed with carbomycin (file name: carbomycin.pdb); large ribosomal subunit complexed with tylosin (file name: tylosin.pdb); large ribosomal subunit complexed with sparsomycin (file name: sparsomycin.pdb); large ribosomal subunit complexed with virginiamycin (file name: virginiamycin.pdb); or large ribosomal subunit complexed with spiramycin (file name: spiramycin.pdb).

In another embodiment, the invention provides a method of obtaining an electron density map of a ribosomal subunit of interest that is only slightly different from the ribosomal subunit whose structure has already been determined, for example, by X-ray crystallography. The method comprises the steps of: (a) producing a crystal of the ribosomal subunit of interest, wherein the crystal is isomorphous; (b) obtaining diffraction amplitudes of the crystal produced in step (a); (c) combining the phases of the crystal of the ribosomal subunit whose structure is already known with the diffraction amplitudes obtained in step (b) to produce a combined data

set; and (d) obtaining an electron density map of the selected ribosomal subunit based on the combined data set obtained in step (c).

In another embodiment, the invention further provides a method of obtaining an electron density map of a ribosomal subunit of interest which is related to a ribosomal subunit whose structure is known. The method comprises the steps of: (a) producing a crystal of a ribosomal subunit of interest, wherein the crystal crystallizes in a different unit cell with different symmetry than the crystal of the ribosomal subunit whose structure is known; (b) obtaining X-ray diffraction data of the crystal of interest; (c) inserting the atomic co-ordinates of the known ribosomal subunit into the unit cell of the crystal of interest and modeling the co-ordinates such that they would be capable of producing theoretical X-ray diffraction data that resembles the X-ray diffraction data obtained in step (b); (d) obtaining phases of the crystal of interest from the modeled co-ordinates in step (c); and (e) obtaining an electron density map of the ribosomal subunit of interest from the X-ray diffraction data obtained in step (b) and the phases obtained in step (d).

In addition, the invention provides a method of obtaining a model of a ribosomal subunit of interest, where the ribosomal subunit of interest diverges significantly from but is still homologous to the ribosomal subunit that was used to generate computed phases. The method comprises the steps of: (a) providing the atomic co-ordinates of the ribosomal subunit whose structure is known; and (b) using homology modeling to produce atomic co-ordinates of the ribosomal subunit of interest.

In another aspect, the invention provides a method of growing a crystal of a ribosome or a ribosomal subunit, as well as crystals resulting from such a method. The method comprises the steps of: (a) isolating a ribosome or a ribosomal subunit; (b) precipitating the ribosome or ribosomal subunit; (c) back-extracting the precipitated ribosome or ribosomal subunit to obtain a solution; (d) seeding the back-extracted solution; (e) growing a crystal of the ribosome or ribosomal subunit from the seeded solution by vapor diffusion at room temperature; and (f) harvesting the crystal. Optionally, the method may further comprise one or more of the following steps: (g) stabilizing the crystal by gradual transfer into a solution containing high salt concentration, for example, from about 1.2 M salt to about 1.7 M salt; (h) maintaining the crystal under such a high salt concentration; and (i) flash freezing the crystal.

In another aspect, the invention provides a method of obtaining X-ray diffraction data for a crystal of a ribosome or a ribosomal subunit. The method comprises the steps of: (a) obtaining a crystal of a ribosome or a ribosomal subunit, wherein the crystal has one or more of the following characteristics (1) an average thickness of greater than 15 μm, and (2) untwinned; and (b) using X-ray crystallography to obtain X-ray diffraction data for the crystal of the ribosome or ribosomal subunit. The present invention also discloses a method of obtaining an electron density map of a ribosome or a ribosomal subunit comprising using the X-ray diffraction data described herein to obtain an electron density map of the ribosome or ribosomal subunit.

In another aspect, the invention provides a method of obtaining X-ray diffraction data for a complex of a ribosome and a ligand, for example, a protein synthesis inhibitor, or a complex of a ribosomal subunit and a ligand. The method comprises the steps of (a) obtaining a crystal of a ribosome or a ribosomal subunit, wherein the crystal has one or more of the following characteristics: (1) an average thickness of greater than 15 μm, and (2) untwinned; (b) diffusing a ligand into the crystal and permitting the ligand to attach to the crystal so as to form a complex; and (c) using X-ray crystallography to obtain X-ray diffraction data for the complex. In an alternative aspect, the invention provides a method of obtaining X-ray diffraction data for a complex of a ribosome and a ligand, for example, a protein synthesis inhibitor or for a ribosomal subunit and a ligand. The method comprises the steps of: (a) obtaining a co-crystal for a complex of a ribosome and a ligand or for a complex of a ribosomal subunit and a ligand, wherein the co-crystal has one or more of the following characteristics: (1) an average thickness of greater than 15 μm, and (2) untwinned; and (b) using X-ray crystallography to obtain X-ray diffraction data for the complex. In either method, the X-ray diffraction data can be used to produce an electron density map for a complex of a ribosome and a ligand or for a complex of a ribosomal subunit and a ligand.

In a preferred embodiment, the invention provides a method of locating the attachment of such a ligand to a ribosome or the attachment of the ligand to a ribosomal subunit. The method comprises the steps of: (a) obtaining X-ray diffraction data for a ribosome or for a ribosomal subunit; (b) obtaining X-ray diffraction data for a complex of a ribosome and a ligand or for a complex of a ribosomal subunit and a ligand; (c) subtracting the X-ray diffraction data obtained in step (a) from the X-ray diffraction data obtained in step (b) to obtain the difference in the X-

ray diffraction data; (d) obtaining phases that correspond to X-ray diffraction data obtained in step (a) using one or more of the techniques selected from the group consisting of MIR, MIRAS, SAD and computation from an existing atomic structure; (e) utilizing the phases obtained in step (d) and the difference in the X-ray diffraction data obtained in step (c) to compute a difference Fourier image of the ligand; and (f) locating the attachment of the ligand to a ribosome or the attachment of the ligand to a ribosomal subunit based on the computations obtained in step (e).

In another embodiment, the invention provides an alternative method of obtaining a map of such a ligand attached to a ribosome or of a ligand attached to a ribosomal subunit. The method comprises the steps of: (a) obtaining X-ray diffraction data for a ribosome or for a ribosomal subunit; (b) obtaining X-ray diffraction data for a complex of a ribosome and a ligand or a complex of a ribosomal subunit and a ligand; (c) obtaining phases that correspond to X-ray diffraction data obtained in step (a) using one or more of the techniques selected from the group consisting of MIR, MIRAS, SAD and computation from an existing atomic structure; and (d) utilizing the phases obtained in step (c) and the X-ray diffraction data obtained in step (b) to compute a map of the ligand and the ribosome or of the ligand and the ribosomal subunit.

In another aspect, the invention provides a computer system comprising: (a) a memory having stored therein data indicative of atomic co-ordinates derived from an electron density map having a resolution of at least about 4.5 Å, more preferably of at least about 3.0 Å, and most preferably of about 2.4 Å and defining a ribofunctional locus of a large subunit of a ribosome; and (b) a processor in electrical communication with the memory, the processor comprising a program for generating a three-dimensional model representative of the ribofunctional locus. In a preferred embodiment, the computer system further comprises a device, for example, a computer monitor, or terminal for providing a visual representation of the molecular model. In another preferred embodiment, the processor further comprises one or more programs to facilitate rational drug design.

In a preferred embodiment, the computer system further comprises at least a portion of the atomic co-ordinates deposited at the Protein Data Bank under accession number PDB ID: 1FFK, 1FFZ, 1FG0, or 1JJ2. In another preferred embodiment, the atomic co-ordinates further define at least a portion of a protein synthesis inhibitor, for example, an antibiotic, more specifically an antibiotic selected from the group consisting of anisomycin, blasticidin,

carbomycin, sparsomycin, spiramycin, tylosin and virginiamycin, complexed with a ribofunctional locus, for example, at least a portion of the atomic co-ordinates recorded on compact disk Disk No. 3 of 3, included herein.

In a preferred embodiment, the ribofunctional locus comprises at least a portion of an active site in the ribosomal subunit, for example, at least a portion of one or more of: a peptidyl transferase site (a portion of which may be defined by a plurality of residues set forth in Table 5); an A-site (a portion of which may be defined by a plurality of residues set forth in Table 6); a P-site (a portion of which may be defined by a plurality of residues set forth in Table 7); a polypeptide exit tunnel (a portion of which may be defined by a plurality of residues set forth in Table 8, Table 9 or Table 10); or an antibiotic binding domain (a portion of which may be defined by a plurality of residues set forth in Table 11, Table 12, Table 13, Table 14, Table 15, Table 16 or Table 17). Plurality of residues shall be considered to include at least 3 residues, preferably at least 5 residues, and more preferably at least 10 residues. The ribofunctional locus may be defined by atoms of ribosomal RNA, one or more ribosomal proteins, or a combination of ribosomal RNA and one or more ribosomal proteins.

In another preferred embodiment, the atomic co-ordinates are produced by molecular modeling. Using the atomic co-ordinates provided herein, the skilled artisan may generate models of any ribosome of interest using conventional techniques, for example, conventional homology modeling, and or molecular replacement techniques. In another embodiment, the atomic co-ordinates are produced by homology modeling using at least a portion of the atomic co-ordinates deposited at the Protein Data Bank under accession number PDB ID: 1FFK, 1FFZ, 1FG0, or 1JJ2, or the atomic co-ordinates included in compact disk Disk No. 3 of 3. In another embodiment, the atomic co-ordinates are produced by molecular replacement using at least a portion of the atomic co-ordinates deposited at the Protein Data Bank under accession number PDB ID: 1FFK, 1FFZ, 1FG0, or 1JJ2, or the atomic co-ordinates included in compact disk Disk No. 3 of 3.

In a preferred embodiment, the atomic co-ordinates define residues that are conserved between ribosomes or ribosomal subunits of pathogens, for example, prokaryotic organisms, and, optionally but more preferably, are also absent from ribosomes or ribosomal subunits of a host

organism, for example, a human. In another preferred embodiment, the atomic co-ordinates may define residues that are conserved between ribosomes or ribosomal subunits of prokaryotic organisms, for example, bacteria, and, optionally but more preferably, are also absent from ribosomal subunits of eukaryotes, for example, a mammal, more preferably, a human. This information can be used, for example, via the use of one or more molecular models, to identify targets for rational drug design that may be exploited to develop new molecules, for example, protein synthesis inhibitors, that disrupt protein synthesis in a pathogen, for example, a bacteria, but do not disrupt or otherwise substantially affect protein synthesis in a host organism, for example, a human.

In another aspect, the invention provides a variety of methods for designing, testing and refining new molecules via rational drug design. For example, the invention provides a method that comprises the steps of: (a) providing a model, for example, a molecular model, having a ribofunctional locus of a large subunit of a ribosome, wherein the model is defined by the spatial arrangement of atoms derived from an electron density map having a resolution of at least about 4.5 Å, more preferably to at least about 3.0 Å, and most preferably to about 2.4 Å; and (b) using the model to identify a candidate molecule having a surface complementary to the ribofunctional locus. Preferably, the candidate molecule stereochemically interfits and more preferably binds with the ribofunctional locus of the large subunit of the ribosome.

In a preferred embodiment, the method comprises one or more additional steps of: producing the candidate molecule identified in such a method; determining whether the candidate molecule, when produced, modulates (for example, induces or reduces) ribosomal activity; identifying a modified molecule; producing the modified molecule; determining whether the modified molecule, when produced, modulates ribosomal activity; and producing the modified molecule for use either alone or in combination with a pharmaceutically acceptable carrier. The candidate molecule and/or the modified molecule may be an antibiotic or antibiotic analogue, for example, a macrolide antibiotic or a macrolide analogue.

In a preferred embodiment, the ribofunctional locus used in such a method comprises at least a portion of an active site in the ribosomal subunit. In another preferred embodiment, the ribofunctional locus is defined by at least a portion of one or more of: a peptidyl transferase site

(a portion of which may be defined by a plurality of residues set forth in Table 5); an A-site (a portion of which may be defined by a plurality of residues set forth in Table 6); a P-site (a portion of which may be defined by a plurality of residues set forth in Table 7); a polypeptide exit tunnel (a portion of which may be defined by a plurality of residues set forth in Table 8, Table 9 or Table 10); or an antibiotic binding domain (a portion of which may be defined by a plurality of residues set forth in Table 11, Table 12, Table 13, Table 14, Table 15, Table 16 or Table 17). The ribofunctional locus may be defined by atoms of ribosomal RNA, one or more ribosomal proteins, or a combination of ribosomal RNA and one or more ribosomal proteins.

In another preferred embodiment, the atomic co-ordinates are used to produce a molecular model in an electronic form. The atomic co-ordinates preferably are produced by molecular modeling. In another embodiment, the atomic co-ordinates are produced by homology modeling using at least a portion of the atomic co-ordinates deposited at the Protein Data Bank under accession number PDB ID: 1FFK, 1FFZ, 1FG0, or 1JJ2, or the atomic co-ordinates included in compact disk Disk No. 3 of 3. In another embodiment, the atomic co-ordinates are produced by molecular replacement using at least a portion of the atomic co-ordinates deposited at the Protein Data Bank under accession number PDB ID: 1FFK, 1FFZ, 1FG0, or 1JJ2, or the atomic co-ordinates included in compact disk Disk No. 3 of 3.

In a preferred embodiment, the atomic co-ordinates may define residues that are conserved among ribosomes or ribosomal subunits of pathogens, for example, prokaryotic organisms, and, optionally but more preferably, are also absent in ribosomes or ribosomal subunits of a host organism, for example, a human. In another preferred embodiment, the atomic co-ordinates may define residues that are conserved between ribosomes or ribosomal subunits of prokaryotic organisms, for example, bacteria, and, optionally but more preferably, are also absent from ribosomes or ribosomal subunits of eukaryotes, for example, a mammal, more preferably a human. This information can be used, for example, via the use of one or more molecular models, to identify targets for rational drug design that may be exploited to develop new molecules, for example, protein synthesis inhibitors, that disrupt protein synthesis in a pathogen, for example, a bacteria but do not disrupt or otherwise substantially affect protein synthesis in a host organism, for example, a human.

In a preferred embodiment, the invention provides a method of obtaining a modified agent. The method comprises the steps of: (a) obtaining a crystal of a ribosome or of a ribosomal subunit; (b) obtaining the atomic co-ordinates of the crystal; (c) using the atomic co-ordinates and one or more molecular modeling techniques, for example, graphic molecular modeling and computational chemistry, to determine how to modify the interaction of an agent with a ribosome or ribosomal subunit; and (d) modifying the agent based on the determinations obtained in step (c) to produce a modified agent. Alternatively, the method further comprises contacting the modified agent with a ribosome or ribosomal subunit and detecting the interaction of the agent to the ribosome or ribosomal subunit. The present invention also provides such a modified agent (preferably a therapeutic agent), wherein the modified agent binds differently to a ribosome or ribosomal subunit than does the agent from which the modified agent was derived.

In another aspect, the invention provides new protein synthesis inhibitors that disrupt the function of a target ribosome. These inhibitors can be readily designed and tested as disclosed herein.

One type of protein synthesis inhibitor of the invention comprises: a first binding domain having a surface, for example, a solvent accessible surface, that mimics or duplicates a surface of a known first molecule, for example, a first antibiotic, that binds with a first contact site, for example, a first ribofunctional locus, in or on a large ribosomal subunit; and a second binding domain having a surface, for example, a solvent accessible surface, that mimics or duplicates a surface of a known second molecule, for example, a second antibiotic, that binds with a second contact site, for example, a second ribofunctional locus, in or on the ribosomal subunit. The first domain is attached to the second domain so as to permit both the first domain and the second domain to bind simultaneously with their respective contact sites within or on the ribosomal subunit so as to disrupt protein synthesis in a ribosomal subunit.

Another type of protein synthesis inhibitor is a synthetic, engineered molecule that comprises: a binding domain having a surface, for example, a solvent accessible surface, that mimics or duplicates a solvent accessible surface of a known molecule, for example, a first known antibiotic, which binds with a contact site, for example, a ribofunctional locus in or on a ribosomal subunit; and an effector domain attached to the binding domain which, upon binding

of the binding domain with the contact site, occupies a space within or adjacent the ribosomal subunit thereby to disrupt protein synthesis in the ribosomal subunit.

The foregoing aspects and embodiments of the invention may be more fully understood by reference to the following figures, detailed description and claims. Further advantages are evident from the drawings (provided in both grayscale and color).

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Color renditions similar to many of the following figures can be found, for example, in Ban *et al.* (2000) *Science* 289: 905-920; or Nissen *et al.* (2000) *Science* 289: 920-929.

The objects and features of the invention may be more fully understood by reference to the drawings described below:

Figures 1(A)-(E) show the electron density from a 2.4 Å resolution electron density map. Specifically, **Figure 1(A)** shows a stereo view of a junction between 23S rRNA domains II, III, and IV. **Figure 1(B)** shows the extended region of protein L2 interacting with surrounding RNA. **Figure 1(C)** shows in detail the L2 region with a bound Mg^{2+} ion. **Figure 1(D)** shows in detail L2 with amino acid side chains. **Figure 1(E)** shows helices 94-97 from domain 6.

Figure 2 shows the *H. marismortui* large ribosomal subunit in the crown view. The subunit is shown in the crown view, with its L7/L12 stalk to the right, its L1 stalk to the left, and its central protuberance (CP) up. In this view, the surface of the subunit that interacts with the small ribosomal subunit faces the reader. RNA is shown in gray in a space-filling rendering. The backbones of the proteins visible are rendered in gold. A transition state analogue bound to the peptidyl transferase site of the subunit is indicated in green. The particle is approximately 250 Å across.

Figures 3(A)-(B) show the secondary structure of the 23S rRNA from *H. marismortui*. The secondary structure of this 23S rRNA is shown in a standardized format. **Figure 3(A)**

shows the 5' half of the large subunit rRNA. **Figure 3(B)** show the 3' half of the large subunit rRNA. This diagram shows all the base pairings seen in the crystal structure of the large subunit that are stabilized by at least two hydrogen bonds. Pairings shown in red were predicted and are observed. Those shown in green were predicted, but are not observed. Interactions shown in blue are observed, but were not predicted. Bases shown in black do not appear to be involved in pairing interactions. Sequences that cannot be visualized in the 2.4 Å resolution electron density map are depicted in gray with the secondary structures predicted for them.

Figures 4(A)-(L) show the tertiary structures of the RNA domains in the *H. marismortui* large ribosomal subunit, its RNA as a whole, and schematics of its RNAs. Specifically, **Figures 4(A)** and **4(B)** show the RNA structure of the entire subunit. Domains are color coded as shown in the schematic of Figure 5(C). **Figure 4(A)** shows the particle in the crown view. **Figure 4(B)** shows the image in Figure 4(A) rotated 180° about an axis running vertically in the plane of the image. **Figures 4(C)** and **4(D)** show a schematic diagram of 23S rRNA and the secondary structure of 5S rRNA. **Figure 4(C)** shows a schematic diagram of 23S rRNA secondary structure of Figure 3 with helices numbered according to Leffers *et al.* ((1987) *J. Mol. Biol.* 195: 43-61), and the domains of the molecule are indicated by color shading. **Figure 4(D)** shows the secondary structure of 5S rRNA from *H. marismortui*. Thick lines joining bases represent Watson-Crick pairing. Bases joined by a lower case "o" indicate non-Watson-Crick pairing. Bases joined by thin lines interact via a single hydrogen bond. Bases shown in black are not paired. Bases shown in red are phylogenetically predicted pairing that have now been confirmed (Symanski *et al.* (1998) *Nucl. Acids Res.* 26: 156-159). Pairs shown in blue are observed, but were not predicted, and pairs shown in green were predicted but are not observed. **Figures 4(E) through 4(L)** show stereo views of the RNA domains in the 23S rRNA and of 5S rRNA. Each domain is color-coded from its 5' end to its 3' end to facilitate the viewer following its trajectory in three-dimensions. The surfaces where the most important inter-domain interactions occur are shown in mono to the right of the stereo views. **Figure 4(E)** shows domain I; **Figure 4(F)** shows domain II; **Figure 4(G)** shows domain III; **Figure 4(H)** shows domain IV; **Figure 4(I)** shows domain V, crown view; **Figure 4(J)** shows domain V, back view; **Figure 4(K)** shows domain VI; and **Figure 4(L)** shows 5S rRNA.

Figures 5(A)-(C) show conservations and expansions in the 23S rRNA of *H. marismortui*. The generality of the RNA in these images is gray. Sequences that are found to be >95% conserved across the three phylogenetic kingdoms are shown in red. Sequences where expansion in the basic 23S structure is permitted are shown in green (Gutell *et al.* (2000) *supra*). Specifically, **Figure 5(A)** shows the particle rotated with respect to the crown view so that its active site cleft can be seen. **Figure 5(B)** shows the crown view. **Figure 5(C)** shows the back view of the particle, *i.e.,* the crown view rotated 180° about its vertical axis.

Figures 6(A)-(I) show structures of some large subunit ribosomal proteins that have non-globular extensions. Only the backbones of the proteins are shown. The globular domains of these proteins are shown in green, and their non-globular extensions are depicted in red. The positions of the zinc ions in L44e and L37e are also indicated. **Figure 6(A)** shows L2; **Figure 6(B)** shows L3; **Figure 6(C)** shows L39; **Figure 6(D)** shows L4; **Figure 6(E)** shows L15; **Figure 6(F)** shows L21e; **Figure 6(G)** shows L44e; **Figure 6(H)** shows L37e; **Figure 6(I)** shows L19;

Figures 7(A)-(C) show proteins that appear on the surface of the large ribosomal subunit. The RNA of the subunit is shown in gray, as in Figure 2, and protein backbones are shown in gold. Specifically, **Figure 7(A)** shows the subunit in the crown view of the subunit. **Figure 7(B)** shows the back side of the subunit in the crown view orientation. **Figure 7(C)** shows the bottom view; the end of the peptide tunnel appears in the center of this image. The proteins visible in each image are identified in the small images at the lower left corner of the Figure.

Figures 8(A)-(F) show the protein distribution and protein-RNA interactions in the large ribosomal subunit. Specifically, **Figure 8(A)** shows the structures of proteins in the neighborhood of the end of the peptide tunnel and how they relate to the RNA sequences with which they interact. Protein L22 extends a long β hairpin extension inside the 23S rRNA. L24 has a similar extension but the entire protein is on the surface of the particle. L39 is the only protein in the subunit that lacks tertiary structure, while L37e has both NH_2 and C00H terminal extensions. L19 is unique in having two globular domains on the surface of the subunit connected by an extended sequence that weaves through the RNA. The end of L39 (green) actually enters the tunnel, while L37e (red) is entirely surrounded by RNA. **Figure 8(B)** shows the non-globular extensions of L2 and L3 reaching through the mass of 23S rRNA towards the

peptidyl transferase site, which is marked by a CCdA-p-puromycin molecule. **Figure 8(C)** shows L22 interacting with portions of all six of the domains of 23S rRNA. **Figure 8(D)** shows a schematic of 23S rRNA showing the locations of the sequences that make at least van der Waals contact with protein (red). **Figure 8(E)** shows a stereo view of the proteins of the large ribosomal subunit with all the RNA stripped away. Proteins are color red as an aid to visualization only. **Figure 8(F)** shows a cross section of the subunit in the area of the tunnel exit. Protein L22 is shown as ribbons in red, and the β hairpin loop where mutations confer erythromycin resistance is shown in orange. Atoms on the surface are shown in gray, protein atoms are shown in green, and atoms at the slice interface are shown in blue.

Figures 9(A)-(C) show chemical structures of ribosome peptidyl transferase substrates and analogues. Specifically, **Figure 9(A)** shows the tetrahedral carbon intermediate produced during peptide bond formation; the tetrahedral carbon is indicated by an arrow. **Figure 9(B)** shows the transition state analogue formed by coupling the 3' OH of CCdA to the amino group of the O-methyl tyrosine residue of puromycin via a phosphate group, CCdA-p-Puro (Welch *et al.* (1995) *supra*). **Figure 9(C)** shows an amino-N-acylated mini helix constructed to target the A-site. The oligonucleotide sequence 5' phosphate CCGGCGGGCUGGUUCAAACCGGCCCGCCGGACC 3' (SEQ ID NO: 1) puromycin should form 12 base pairs. The construct was based on a mini helix which is a suitable substrate for amino-acylation by Tyr-tRNA synthetase. The 3' OH of its terminal C is coupled to the 5' OH of the N6-dimethyl A moiety of puromycin by a phosphodiester bond.

Figures 10(A)-(C) show experimentally phased electron density maps of the substrate analogue complexes at 3.2 Å resolution, with models superimposed (oxygen, red; phosphorus, purple; nitrogen, blue; and carbon, green for rRNA and yellow for substrate). Specifically, **Figure 10(A)** shows an F_o(complex)- F_o(parent) difference electron density map with a skeletal model of CCdA-p-Puro superimposed. **Figure 10(B)** shows a $2F_o$(complex)- F_o(parent) electron density map of the CCdA-p-Puro in the active site region with the structures of the ribosome and inhibitor superimposed showing the proximity of the N3 of A2486 (2451) to the phosphate, non-bridging oxygen in this complex. **Figure 10(C)** shows an F_o(complex)- F_o(parent) differences electron density map of the tRNA acceptor stem analogue with a skeletal model of CCpuro

superimposed. There is density only for the ribose and phosphate of C74 and none for the rest of the RNA hairpin.

Figures 11(A) and (B) show a combined model of the CCA portion of the mini helix bound to the A-site and CCdA-p-Puro bound to the A- and P-sites, color coded as in Figure 2. Specifically, **Figure 11(A)** shows the base-pairing interactions between the P-site C74 and C75 and the P loop of 23S rRNA on the left and the A-site C75 with the A loop of 23S rRNA on the right. The catalytic A2486 is near the phosphate oxygen (P) that is the analogue of the tetrahedral intermediate oxyanion. **Figure 11(B)** shows A2637 (in all blue) lying between the two CCA's and A2486 (green) whose N3 approaches a non-bridging phosphate oxygen. The N1 atoms of the A76 bases from the A- and P-site tRNAs are making nearly identical interactions with a ribose 2' OH in both the A- and P-loops, respectively, and an approximate 2-fold axis relates these residues.

Figure 12 shows a space filling model of the 23S and 5S rRNA, the proteins and the CCdA-p-Puro inhibitor viewed down the active site cleft in a rotated "crown view." The bases are white and the sugar phosphate backbones are yellow. The inhibitor is shown in red and the numbered proteins are shown in blue. The L1 and L11 proteins positioned at lower resolution are in blue backbone. The central protuberance is labeled CP.

Figure 13(A) shows a stereo view diagram of the three-dimensional distribution of the residues comprising the loops A and P and the peptidyl transferase loop. **Figure 13(B)** shows a stereo view of the central loop in domain V from the direction of the tunnel. The residues are color coded based on mutations which confer antibiotic resistance. **Figure 13(C)** shows domain V active site with its central loop shown as the secondary structure.

Figures 14(A) and (B) show the closest approach of polypeptides to the peptidyl transferase active site marked by a ball and stick representation of the Yarus inhibitor, CCdA-p-Puro. Specifically, **Figure 14(A)** shows a coil representation of domain V RNA backbone in red and bases in gray and a ribbon backbone representation of all thirteen proteins that interact with it. **Figure 14(B)** shows a close-up view of the active site with the RNA removed. The phosphate of the Yarus analogue and the proteins whose extensions are closest to the inhibitor are shown in ribbon with their closest side-chains in all atom representation. The distances in Å

between the closest protein atoms and the phosphorous analogue of the tetrahedral carbon (pink) are shown, as is a modeled peptide (pink).

Figure 15 shows conserved nucleotides in the peptidyl transferase region that binds CCdA-p-Puro A space filling representation of the active site region with the Yarus inhibitor viewed down the active site cleft. All atoms belonging to 23S rRNA nucleotides that are 95% conserved in all three kingdoms (Gutell *et al.* (2000) *supra*) are colored red and all other nucleotides are white; the inhibitor is colored blue.

Figures 16(A)-(C) show the catalytic apparatus of the peptidyl transferase active site. Specifically, **Figure16(A)** shows stereo view of a portion of the experimental 2.4 Å resolution electron density map (Ban *et al.* (2000) *Science* 289: 905-920) of the large subunit in the region of the catalytic site in stereo. The structure the RNA involved in interactions with A2486 is superimposed. Residues G2102 (2061) and G2482 (2447) are hydrogen bonded to the N6 of A2486 (2451) and G2482 which interacts with a neighboring phosphate group. **Figure 16(B)** shows a skeletal representation with dashed hydrogen-bonds showing G2482, G2102, A2486 and the buried phosphate that is proposed to result in a charge relay through G2482 to the N3 of A2486. **Figure 16(C)** shows the normal and rarer imine tautomeric forms of G2482 and A2486 that are proposed to be stabilized by the buried phosphate of residue 2485.

Figures 17(A)-(C) show the proposed mechanism of peptide synthesis catalyzed by the ribosome. Specifically, **Figure17(A)** shows the N3 of A2486 abstracting a proton from the NH_2 group as the latter attacks the carbonyl carbon of the peptidyl-tRNA. **Figure 17(B)** shows a protonated N3 stabilizing the tetrahedral carbon intermediate by hydrogen bonding to the oxyanion. **Figure 17(C)** shows the proton transferred from the N3 to the peptidyl tRNA 3' OH as the newly formed peptide deacylates.

Figures 18(A) and (B) show space filling representations of the 50S ribosomal subunit with the 3 tRNA molecules, in the same relative orientation that they are found in the 70S ribosome structure by Noller and colleagues docked onto the CCA's bound in the A-Site and P-Site. Specifically, **Figure 18(A)**, shown on the left-hand side, shows the whole subunit in rotated crown view with the rRNA in yellow, proteins in pink and tRNAs in orange. **Figure 18(B)**, shown on the right-hand side, shows a close-up view showing the numbered proteins are

in pink and the rRNA in blue. A backbone ribbon representation of the A-, P-, and E-sites are shown in yellow, red and white, respectively.

Figures 19(A)-(F) show the polypeptide exit tunnel. Specifically, **Figure 19(A)** shows the subunit cut in half, roughly bisecting its central protuberance and its peptide tunnel along the entire length. The two halves have been opened like the pages of a book. All ribosome atoms are shown in CPK representation, with all RNA atoms that do not contact solvent shown in white and all protein atoms that do not contact solvent shown in green. Surface atoms of both protein and RNA are color-coded with carbon in yellow, oxygen in red, and nitrogen in blue. A possible trajectory for a polypeptide passing through the tunnel is shown as a white ribbon. The peptidyl transferase site (PT) is also shown. **Figure 19(B)** shows detail of the polypeptide exit tunnel with the distribution of polar and non-polar groups, with atoms colored as in Figure 19(A), the constriction in the tunnel formed by proteins L22 and L4 (green patches close to PT), and the relatively wide exit of the tunnel. A modeled polypeptide is in white. **Figure 19(C)** shows the tunnel surface with backbone atoms of the RNA color coded by domain: domain I (white), II (light blue), III (gold), IV (green), V (orange), 5S (pink) and proteins are blue. The peptidyl transferase center (PTC) is shown. **Figure 19(D)** is a space filling representation of the large subunit surface at the tunnel exit showing the arrangement of proteins, some of which might play roles in protein secretion. The RNA is in white (bases) and yellow (backbone) and the numbered proteins are blue. A modeled polypeptide is exiting the tunnel in red. **Figure 19(E)** shows a close-up view of the half of the exit tunnel showing the relationship of the peptidyl transferase center (PTC) to proteins L4 (yellow) and L22 (blue). The Yarus inhibitor and a modeled peptide are purple and the 23S rRNA is in red and white. **Figure 19(F)** shows a secondary structure schematic of 23S rRNA identifying the sequences that contact the tunnel in red.

Figure 20 is a picture showing the spatial relationship between the antibiotic anisomycin bound to a large ribosomal subunit.

Figure 21 is a picture showing the spatial relationship between the antibiotic blasticidin bound to a large ribosomal subunit.

Figure 22 is a picture showing the spatial relationship between the antibiotics carbomycin and tylosin bound to a large ribosomal subunit.

Figure 23 is a picture showing the spatial relationship between the antibiotic sparsomycin bound to a large ribosomal subunit.

Figure 24 is a picture showing the spatial relationship between the antibiotics virginiamycin (streptogramin A) and carbomycin bound to a large ribosomal subunit.

Figure 25 is a picture showing the spatial relationship of certain antibiotics, namely, anisomycin, blasticidin, carbomycin, and virginiamycin, bound to a large ribosomal subunit. The locations of the bound antibiotics are shown relative to the ribosomal A-site, P-site, and polypeptide exit tunnel.

Figures 26(A)-(C) are pictures showing a peptidyl transferase site disposed within a large ribosomal subunit. **Figure 26A** shows a bound tylosin molecule, and identifies a disaccharide binding pocket and two cavities denoted "cavity 1" and "cavity 2." **Figures 26(B) and (C)** are provided on the left hand side to orient the reader to the locations of the peptidyl transferase site (PT) and polypeptide exit tunnel in the large ribosomal subunit.

Figure 27 is a schematic representation of a computer system useful in molecular modeling a ribosomal subunit and/or for performing rational drug design.

Figure 28 is a schematic representation of certain potential drug target sites in a large ribosomal subunit.

Figures 29(A)-(D) are pictures showing the residues within the wall of the polypeptide exit tunnel that are conserved (red) or non-conserved (blue) between *E. coli* and rat. The ribosomal subunit has been sliced down the polypeptide exit tunnel with one half of the polypeptide exit tunnel shown in **Figure 29(A)**, and the other half of the polypeptide exit tunnel is shown in **Figure 29(B)**. **Figure 29(C)** is provided to orient the reader to show the location of the portion of the ribosomal subunit shown in Figure 29(A) relative to the ribosomal subunit as a whole. **Figure 29(D)** is provided to orient the reader to show the location of the portion of the ribosomal subunit shown in Figure 29(B) relative to the large ribosomal subunit as a whole.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

As used herein, the term "active site" refers to regions on a ribosome or ribosomal subunit that are directly involved in protein synthesis, *e.g.*, the peptidyl transferase site, the elongation factor binding site, and other similar sites.

As used herein, the terms "agent" and "ligand" are used synonymously and refer to any atom, molecule, or chemical group which binds with a ribosome, ribosomal subunit or ribosome fragment. Thus, ligands include, but are not limited to, a single heavy atom, an antibiotic, a tRNA, a peptidyl tRNA, an aminoacyl tRNA, or a signal recognition particle ("SRP").

As used herein, "archaebacteria" refers to the kingdom of monerans that includes methane producers, sulfur-dependent species, and many species that tolerate very salty or hot environments.

As used herein, the term "A-site" refers to the locus occupied by an aminoacyl-tRNA molecule immediately prior to its participation in the peptide-bond forming reaction.

As used herein, the term "asymmetric unit" refers to a minimal set of atomic co-ordinates that when operated upon by the symmetry operations of a crystal will regenerate the entire crystal.

As used herein, "at least a portion of" or "at least a portion of the three-dimensional structure of" is understood to mean a portion of the three-dimensional structure of a ribosome or ribosomal subunit, including charge distribution and hydrophilicity/hydrophobicity characteristics, formed by at least three, more preferably at least three to ten, and most preferably at least ten contiguous amino acid and/or nucleotide residues of the ribosome or ribosomal subunit. The contiguous residues forming such a portion may be residues which form a contiguous portion of the primary sequence of a ribosomal RNA or ribosomal protein, residues which form a contiguous portion of the three-dimensional structure of the ribosome or ribosomal subunit, or a combination thereof. Thus, the residues forming a portion of the three-dimensional structure need not be contiguous in the primary sequence but, rather, must be contiguous in space. As used herein, the residues forming "a portion of the three-dimensional structure of" a

ribosome or ribosomal subunit, form a contiguous three-dimensional shape in which each atom or functional group forming the portion of the shape is separated from the nearest atom or functional group forming the portion of the shape by no more than 40 Å, preferably by no more than 20 Å, more preferably by no more than 5-10 Å, and most preferably by no more than 1-5 Å.

As used herein, the term "atomic co-ordinates" or "structure co-ordinates" refers to mathematical co-ordinates (represented as "X," "Y" and "Z" values) that describe the positions of atoms in a crystal of a ribosome or ribosomal subunit. The diffraction data obtained from the crystals are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are used to establish the positions of the individual atoms within a single ribosomal subunit. Those of skill in the art understand that a set of structure co-ordinates determined by X-ray crystallography is not without standard error. For the purpose of this invention, any set of structure co-ordinates for a ribosome or ribosomal subunit from any source has a root mean square deviation of non-hydrogen atoms of less than 0.75 Å when superimposed on the non-hydrogen atom positions of the said atomic co-ordinates deposited at the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Berman *et al.* (2000) *Nucleic Acids Research* 28, 235-242; http://www.rcsb.org/pdb/) with the accession numbers PDB ID: 1FFK; PDB ID: 1FFZ; PDB ID: 1FG0; or PDB ID: 1JJ2, the disclosure of each of the foregoing of which is incorporated herein by reference in its entirety.

In the list of atomic co-ordinates deposited at the RCSB Protein Data Bank or included herein as files recorded on the compact disks, the term "atomic co-ordinate" or structure co-ordinates refer to the measured position of an atom in the structure in Protein Data Bank (PDB) format, including X, Y, Z and B, for each. The term "atom type" refers to the element whose co-ordinates are measured. The first letter in the column defines the element. The term "X", "Y", "Z" refers to the crystallographically defined atomic position of the element measured with respect to the chosen crystallographic origin. The term "B" refers to a thermal factor that measures the mean variation of an atom's position with respect to its average position.

Reference is made to the sets of atomic co-ordinates and related tables included with this specification and submitted on compact disk (six total compact disks including three original compact disks, and a duplicative copy of each of the original compact disks), all of the foregoing of which are incorporated by reference herein. Disk No. 1 of 3 contains eight files; Disk No. 2 of

3 contains four files; and Disk No. 3 of 3 contains nine files. Disk No. 1 of 3 contains the files identified as PDB1FFK.DOC and PDB1FFK.ENT which represent files of co-ordinates defining the large ribosomal subunit; PDB1FFZ.DOC and PDB1FFZ.ENT which represent files of the co-co-ordinates defining the large ribosomal subunit - CCdA-p-Puro complex; and PDB1FGO.DOC and PDB1FGO.ENT which represent files of the co-ordinates defining the large ribosomal subunit - aa-tRNA analogue complex. Disk No. 2 of 3 contains files identified as 1JJ2.RTF and 1JJ2.TXT which represent files of the co-ordinates defining the completely refined large ribosomal subunit. Disk No. 3 of 3 contains the files identified as anisomycin.pdb, blasticidin.pdb, carbomycin.pdb, sparsomycin.pdb, spiramycin.pdb, tylosin.pdb and virginiamycin.pdb which represent files of the co-ordinates defining the large ribosomal subunit bound to anisomycin, blasticidin, carbomycin, sparsomycin, spiramycin, tylosin, and virginiamycin, respectively.

As will be apparent to those of ordinary skill in the art, the atomic structures presented herein are independent of their orientation, and that the atomic co-ordinates identified herein merely represent one possible orientation of a particular large ribosomal subunit. It is apparent, therefore, that the atomic co-ordinates identified herein may be mathematically rotated, translated, scaled, or a combination thereof, without changing the relative positions of atoms or features of the respective structure. Such mathematical manipulations are intended to be embraced herein.

As used herein, the terms "atomic co-ordinates derived from" and "atoms derived from" refers to atomic co-ordinates or atoms derived, either directly or indirectly, from an electron density map. It is understood that atomic co-ordinates or atoms derived "directly" from an electron density map refers to atomic co-ordinates or atoms that are identified from and/or fitted to an electron density map by using conventional crystallographic and/or molecular modeling techniques and thus can be considered to be primary atomic co-ordinates or atoms. It is understood that atomic co-ordinates or atoms derived "indirectly" from an electron density map refers to atomic co-ordinates or atoms that are derived from and thus are derivatives or transforms of the primary atomic co-ordinates or atoms and thus can be considered to be secondary atomic co-ordinates or atoms. The secondary atomic co-ordinates or atoms may be generated from the primary atomic co-ordinates or atoms by using conventional molecular

modeling techniques. By way of a non limiting example, the atomic co-ordinates for the *H. marismortui* large ribosomal subunit as described hereinbelow are considered to be primary co-ordinates, whereas the atomic co-ordinates of a mammalian large ribosomal subunit which can be derived from *H. marismortui* atomic co-ordinates by molecular modeling, including, for example, homology modeling and/or molecular replacement, are considered to be secondary co-ordinates. Both types of atomic co-ordinates and atoms are considered to be embraced by the invention.

As used herein the terms "bind," "binding," "bound," "bond," or "bonded," when used in reference to the association of atoms, molecules, or chemical groups, refer to any physical contact or association of two or more atoms, molecules, or chemical groups (*e.g.,* the binding of a ligand with a ribosomal subunit refers to the physical contact between the ligand and the ribosomal subunit). Such contacts and associations include covalent and non-covalent types of interactions.

As used herein, the terms "complex" or "complexed" refer to the assembly of two or more molecules to yield a higher order structure, such as, a 50S ribosomal subunit bound to a ligand.

As used herein, the term "computational chemistry" refers to calculations of the physical and chemical properties of the molecules.

As used herein, the term "conjugated system" refers to more than two double bonds that are positioned spatially so that their electrons are completely delocalized with the entire system. Aromatic residues contain conjugated double bond systems.

As used herein, the terms "covalent bond" or "valence bond" refer to a chemical bond between two atoms in a molecule created by the sharing of electrons, usually in pairs, by the bonded atoms.

As used herein, the term "crystal" refers to any three-dimensional ordered array of molecules that diffracts X-rays.

As used herein, the term "crystallographic origin" refers to a reference point in the unit cell with respect to the crystallographic symmetry operation.

As used herein, the term "elongation factor binding domain" refers to the region of the ribosome that interacts directly with elongation factors, including, for example, the elongation factors, EF-Tu and EF-G.

As used herein, the term "E-site" refers to the locus occupied by a deacylated tRNA molecule as it leaves the ribosome following its participation in peptide-bond formation.

As used herein, the term "heavy atom derivatization" refers to the method of producing a chemically modified form, also known as a "heavy atom derivative", of a crystal of the ribosome and the ribosomal subunit and its complexes. In practice, a crystal is soaked in a solution containing heavy metal atom salts, or organometallic compounds, *e.g.,* mercury chlorides, ethyl-mercury phosphate, osmium pentamine, or iridium pentamine, which can diffuse through the crystal and bind to the ribosome or ribosomal subunit. The location(s) of the bound heavy metal atom(s) can be determined by X-ray diffraction analysis of the soaked crystal. This information, in turn, is used to generate the phase information used to construct three-dimensional structure of the complex (Blundell *et al.* (1976) *supra*).

As used herein, the term "homologue" is understood to mean any one or combination of (i) any protein isolated or isolatable from a ribosome or a ribosomal subunit (*i.e.,* a ribosomal protein), (ii) any nucleic acid sequence isolated or isolatable from a ribosome or ribosomal subunit (*i.e.,* a ribosomal RNA), (iii) any protein having at least 25 % sequence identity to a ribosomal protein isolated from *E. coli* or *Rattus norvegicus* as determined using the computer program "BLAST" version number 2.1.1 implementing all default parameters, or (iv) any nucleic acid having at least 30% sequence identity to a ribosomal RNA isolated from *E. coli* or *Rattus norvegicus* as determined using the computer program "BLAST" version number 2.1.1 implementing all default parameters. "BLAST" version number 2.1.1 is available and accessible via the world wide web at http://www/ncbi.nlm.nih.gov/BLAST/ or can be run locally as a fully executable program on a standalone computer.

As used herein, the term "homology modeling" refers to the practice of deriving models for three-dimensional structures of macromolecules from existing three-dimensional structures for their homologues. Homology models are obtained using computer programs that make it possible to alter the identity of residues at positions where the sequence of the molecule of interest is not the same as that of the molecule of known structure.

As used herein, the term "hydrogen bond" refers to two electronegative atoms (either O or N), which share a hydrogen that is covalently bonded to only one atom, while interacting with the other.

As used herein, the term "hydrophobic interaction" refers to interactions made by two hydrophobic residues.

As referred to herein, ribosomal proteins are designated "LX" or "SX", where L stands for "large subunit; S stands for "small subunit"; and X in either case is an integer.

As used herein, the term "MIR" refers to multiple isomorphous replacement, a technique used for deriving phase information from crystals treated with heavy atom compounds.

As used herein, the term "molecular graphics" refers to three-dimensional representations of atoms, preferably on a computer screen.

As used herein, the terms "molecular model" or "molecular structure" refer to the three-dimensional arrangement of atoms within a particular object (*e.g.*, the three-dimensional structure of the atoms that comprise a ribosome or ribosomal subunit, and the atoms that comprise a ligand that interacts with a ribosome or ribosomal subunit, particularly with a large ribosomal subunit, more particularly with a 50S ribosomal subunit).

As used herein, the term "molecular modeling" refers to a method or procedure that can be performed with or without a computer to make one or more models, and, optionally, to make predictions about structure activity relationships of ligands. The methods used in molecular modeling range from molecular graphics to computational chemistry.

As used herein, the term "molecular replacement" refers to a method that involves generating a model of a ribosome or ribosomal subunit whose atomic co-ordinates are unknown, by orienting and positioning the atomic co-ordinates described in the present invention in the unit cell of the crystals of the unknown ribosome so as best to account for the observed diffraction pattern of the unknown crystal. Phases can then be calculated from this model and combined with the observed amplitudes to give the atomic co-ordinates of the unknown ribosome or ribosomal subunit. This type of method is described, for example, in The Molecular Replacement Method, (Rossmann, M.G., ed.), Gordon & Breach, New York, (1972).

As used herein, "noncovalent bond" refers to an interaction between atoms and/or molecules that does not involve the formation of a covalent bond between them.

As used herein, the term "peptidyl transferase site" refers to the locus in the large ribosomal subunit where peptide bonds are synthesized.

As used herein, the term "polypeptide exit tunnel" refers to the channel that passes through the large ribosomal subunit from the peptidyl transferase site to the exterior of the ribosome through which newly synthesized polypeptides pass.

As used herein, the term "protein synthesis inhibitor" refers to any molecule that can reduce, inhibit or otherwise disrupt protein or polypeptide synthesis in a ribosome.

As used herein, the term "P-site" refers to the locus occupied by a peptidyl-tRNA at the time it participates in the peptide-bond forming reaction.

As used herein, the term "ribofunctional locus" refers to a region of the ribosome or ribosomal subunit that participates, either actively or passively, in protein or polypeptide synthesis within the ribosome or ribosomal subunit and/or export or translocation of a protein or polypeptide out of a ribosome. The ribofunctional locus can include, for example, a portion of a peptidyl transferase site, an A-site, a P-site, an E-site, an elongation factor binding domain, a polypeptide exit tunnel, and a signal recognition particle (SRP) binding domain. It is understood that the ribofunctional locus will not only have a certain topology but also a particular surface chemistry defined by atoms that, for example, participate in hydrogen bonding (for example, proton donors and/or acceptors), have specific electrostatic properties and/or hydrophilic or hydrophobic character.

As used herein, the term "ribosomal subunit" refers to one of the two subunits of the ribosome that can function independently during the initiation phase of protein synthesis but which both together constitute a ribosome. For example, a prokaryotic ribosome comprises a 50S subunit (large subunit) and a 30S subunit (a small subunit).

As used herein, the term "ribosome" refers to a complex comprising a large ribosomal subunit and small ribosomal subunit.

As used herein, the term "signal recognition particle binding domain" refers to the portion of the ribosome that interacts directly with the signal recognition particle.

As used herein, the term "space group" refers to the arrangement of symmetry elements of a crystal.

As used herein, the term "symmetry operation" refers to an operation in the given space group that places the atoms in one asymmetric unit on the corresponding atoms in another asymmetric unit.

As used herein, the term "twinned" refers to a single macroscopic crystal that contains microscopic domains of the same symmetry that differ significantly in orientation in such a way that the diffraction patterns of all are superimposed. In a twinned crystal the mosaic blocks, or domains, are orientated so that some point in one direction and others point in a second, distinctly different direction, and the directions are such that the diffraction pattern generated by one group of blocks falls exactly on top of the diffraction pattern of the other group.

As used herein, the term "untwinned" refers to a crystal cell the domains of which are aligned. The domains are also known as the "mosaic blocks." Most crystals diffract as though they were assemblies of mosaic blocks. One can think of them as small, perfectly ordered regions within the larger crystal, which, overall, is not so well ordered. Each block has the same symmetry and unit cell packing as all the others.

As used herein, the term "unit cell" refers to a basic parallelepiped shaped block. The entire volume of crystal may be constructed by regular assembly of such blocks. Each unit cell comprises a complete representation of the unit of pattern, the repetition of which builds up the crystal.

II. Structure and Use of the Large Ribosomal Subunit

A. Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution, Initial Refinement

The present invention is based, in part, on the development of a novel method for preparing crystals of ribosomes. The novel method provides crystals of the 50S ribosomal subunit that are much thicker than those available earlier and that can diffract X-rays to a

resolution of about 2.4 Å. The method eliminates the twinning of crystals that obstructed progress in determining the crystal structure of the 50S ribosomal subunit from *H. marismortui* for many years. The method of preparing the crystals of the 50S ribosomal subunit is discussed below.

The present invention is also based, in part, on the atomic structure of the crystal of the 50S ribosomal subunit from *H. marismortui* that has been derived from a 2.4 Å resolution electron density map that was experimentally phased using heavy atom derivatives. The atomic co-ordinates defining the large ribosomal unit were deposited on July 10, 2000, at Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Berman *et al.* (2000) *Nucleic Acid Research* 28, 235-242; http:// www.rcsb.org/pdb/) with accession number PDB ID: 1FFK.

Moreover, the present invention is based, in part, on the derivation from the atomic co-ordinates of the following model which is briefly summarized here and discussed in detail in the following sections of the specification. This model includes 2,811 of the 2,923 nucleotides of 23S rRNA, all 122 nucleotides of its 5S rRNA, and structures for the 27 proteins that are well-ordered in the subunit.

The secondary structures of both 5S and 23S rRNA are remarkably close to those deduced for them by phylogenetic comparison. The secondary structure of the 23S rRNA divides it into 6 large domains, each of which has a highly asymmetric tertiary structure. The irregularities of their shapes notwithstanding, the domains fit together in an interlocking manner to yield a compact mass of RNA that is almost isometric. The proteins are dispersed throughout the structure, concentrated largely on its surface, but they are much less abundant in the regions of the subunit that are of primary functional significance to protein syntheses - the 30S subunit interface, the binding regions for tRNA and the peptidyl transferase active site. The most surprising feature of many of these proteins are the extended, irregular structures of their loops and termini, which penetrate between RNA helices. The primary role of most of the proteins in the subunit appears to be stabilization of the three-dimensional structure of its rRNA.

1. Preparation of the Crystal for the 50S Ribosomal Subunit and Structure Determination.

Several experimental approaches were used to extend the resolution of the electron density maps of the *H. marismortui* 50S ribosomal subunit from 5 Å to 2.4 Å including improvements in the crystals. A back-extraction procedure was developed for reproducibly growing crystals that are much thicker than those available earlier and can diffract to 2.2 Å resolution (see Example 1). Briefly, the crystals were grown at room temperature in hanging drops by vapor diffusion from seeded solutions back-extracted from precipitated subunits. The crystals that resulted had maximum dimensions of 0.5 x 0.5 x 0.2 mm and were harvested after three weeks. The twinning of crystals that obstructed progress for many years (Ban *et al.* (1999) *supra*) was eliminated by adjusting crystal stabilization conditions (see Example 1). Crystals were stabilized by gradual transfer into a solution containing 12%, PEG 6000, 22% ethylene glycol, 1.7 M NaCl, 0.5 M NH_4Cl, 100 mM potassium acetate, 30 mM $MgCl_2$ and 1 mM $CdCl_2$, pH 6.2, and flash frozen in liquid propane. Reducing the salt concentration below 1.7 M NaCl (KCl) increased the tendency of crystals to become twinned. At salt concentrations as low as 1.2 M nearly all of the crystals were twinned.

All the X-ray data used for high resolution phasing were collected at the Brookhaven National Synchrotron Light Source except for two native data sets used, which were collected at the Advanced Photon Source at Argonne (see Example 2) (Table 1). Osmium pentamine (132 sites) and Iridium hexamine (84 sites) derivatives proved to be the most effective in producing both isomorphous replacement and anomalous scattering phase information to 3.2 Å resolution (see Example 2). Inter-crystal density averaging which had contributed significantly at lower resolution, was not helpful beyond about 5 Å resolution. Electron density maps were dramatically improved and their resolutions extended, eventually to 2.4 Å, using the solvent flipping procedure in CNS (Abrahams *et al.* (1996) *Acta Crystollogr.* D 52: 30; Brünger *et al.* (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54: 905-921).

Table 1

Statistics for Data Collection, Phase Determination, and Model Construction

Data Statistics

	MIRAS1			MIRAS2			
	Native1	$Os(NH_3)_5^{2+}$	$UO_2F_5^{3-}$	Native2	$Ir(NH_3)_6^{4+}$	$Os(NH_3)_6^{3+}$	$Ta_6Br_{12}^{2+}$
Heavy atom conc. (mM)	-	30.0	0 5	-	20 0	4 5	3.0
Soaking time (hrs)	-	1 5	4	-	24 hrs	24 hrs	24 hrs
Sites no	-	132	20	-	84	38	9
Resolution (Å) (*)	90-2 4 (2.5-2.4)	40-3 5 (3 6-3 5)	40-3 8 (3.9-3.8)	30-2 9 (3.0-2 9) (3 32-3 22)	30-3 2 (3.27-3 20)	30-3.5 (3 6-3 5)	30-3 8 (3 97-3 80)
λ(Å)	1 00	1 14	1.30	1.00	1.075	1 14	1 255
Observations	6,089,802	1,308,703	596,166	2,832,360	1,823,861	1,646,468	1,288,524
Unique	665,928	429,761	313,863	390,770	541,488	488,275	346,745
Redun. (*)	9 1 (6.5)	3 0 (2 5)	1.9 (1.6)	7 2	3.4	4 3 (4 2)	3 7
Completeness (*)	95 6 (71 0)	99 4 (96 8)	92.0 (54 1)	97 1	93 8	98 1(99 0)	99 5
$I/\sigma I$ (Last bin)	25 5 (1 9)	13 5 (3 3)	8 9 (1 6)	18 0 (6 4)	12 0 (2.6)	10.6 (2 7)	10 8 (3 2)
R_{merge} (*)	8.6 (69 1)	7 2 (32 0)	9 1 (37 9)	11 2 (36 9)	8 5 (29 5)	12.1 (46.0)	12.1 (40.5)
02 (ano) (*)	-	2.8 (1 0)	1 5 (1 0)	-	2.63 (1.48)	1.8 (1 0)	2 42 (1.18)
R_{merge} (ano)	-	6 2	8 0	-	6 7	6 9	
R_{iso}	-	14 1(22 7)	26 4 (47 0)	-	12 9 (28 1)	19 5 (39 4)	

Phasing Statistics

	Resolution shells (Å): ~ 73,200 reflections per bin				
	30.0	**5.1**	**4.0**	**3.5**	**3.2**	**Total**
MIRAS1 (FOM)	0 52	0 31	0 14	-		0 32
$Os(NH_3)_5^{2+}$						
Phasing power	0 87	0.72	0 66	-		0 75
Phasing power (SAD)	1 40	0 58	0 26	-		0 75
R_{cullis} (centric)	0 62	0 65	0 67	-		0 65
$UO_2F_5^{3-}$						
Phasing Power	0 47	0 33	0 28	-		0 36
Phasing power (SAD)	0 46	0 25	-	-		0 36
R_{cullis} (centric)	0 72	0 77	0 75	-		0 75
MIRAS2 (FOM)	0 48	0 40	0 28	0 12		0 33
$Ir(NH_3)_6^{3+}$						
Phasing power	1 02	0 92	0 78	0 66		0 89
Phasing power (SAD)	2 02	1 60	1.22	0 83		1 47
R_{cullis} (centric)	0 58	0 63	0 70	0 74		0 63
$Os(NH_3)_6^{3+}$						
Phasing power	0 62	0 57	0 58	0 58		0 59
Phasing power (SAD)	0 47	0 39	-	-		0 42
R_{cullis} (centric)	0 78	0 78	0 78	0 76		0 78
$Ta_6Br_{12}^{2+}$ (Used for SAD phasing only)						
Phasing Power (SAD)	2 77	0 35	0 13	-		1 19
FOM (MIRAS1+MIRAS2+SAD)	0.76	0 51	0 31	0 14		0 37

Model Statistics

		rms deviations		Average *B* factors (Å^2)	
Resolution range (Å)	90 0-2 4	Bonds (Å)	0.0064	All atoms	37 4
Reflections	577,304	Angles (°)	1 19	23*S* rRNA	32 3
R_{sryst} (%)	25 2	Dihedrals (°)	28 8	5*S* rRNA	43 2
R_{free} (%)	26 1	Impropers (°)	1 68	Minimum/Max *B* factors (Å^2)	70/107 9

λ, wavelength; Redun., redundancy; (*) last-resolution shell.

R_{iso}: $\Sigma|F_{PH}-F_P|/\Sigma F_{PH}$, where F_{PH} and F_P are the derivative and the native structure factor amplitudes, respectively. R_{sym}: $\Sigma\Sigma_i|I_{(h)}-I_{(h)i}|/\Sigma\Sigma:I_{(h)i}$, where I(h) is the mean intensity after reflections. Phasing power: r.m.s. isomorphous difference divided by the r.m.s. residual lack of closure. R_{cullis}: $\Sigma(||F_{PH}-F_P|-|F_{H(calc)}||)/\Sigma|F_{PH}-F_P|$, where F_{PH} is the structure factor of the derivative and F_P is that of the native data. The summation is valid only for centric reflection. FOM (figure of merit): mean value of the cosine of the error in phase angles. Abbreviations: MIRAS: multiple isomorphous replacement, anomalous scattering; SAD: single wavelength anomalous diffraction; FOM: figure of merit

Except for regions obscured by disorder, the experimentally-phased, 2.4 Å resolution electron density map was of sufficient quality so that both protein and nucleic acid sequencing errors could be identified and corrected. Each nucleotide could be fitted individually and the difference between A and G was usually clear without reference to the chemical sequence, as was the distinction between purines and pyrimidines (Figure 1).

Subtraction of the atomic model from the experimental electron density map leaves no significant density except for water and ions, showing that the model accounts for all the macromolecular density. Preliminary refinement of the model was achieved using a mixed target in the program CNS (Brünger *et al.* (1998) *supra*). The model was further refined in real space against the 2.4 Å electron density map using the program TNT (Tronrud (1997), Macromolecular Crystallography, Part B, Methods In Enzymology), which yielded a model with a free R-factor of 0.33. One additional round of mixed target refinement of both atomic positions and B-factors using CNS led to the structure described below. Its free R-factor is 0.27 (Table 1).

2. *Sequence Fitting and Protein Identification.*

The sequence of 23S rRNA was fit into the electron density map nucleotide by nucleotide starting from its sarcin/ricin loop sequence (A2691-A2702) (*E. coli* numbers A2654 to A2665), whose position had been determined at 5 Å resolution (Ban *et al.* (1999) *supra*). Guided by the information available about the secondary structures of 23S rRNAs (Gutell, R.R. (1996), "Comparative Sequence Analysis and the Structure of 16S and 23S rRNA," Ribosomal RNA. Structure, Evolution, Processing, and Function in Protein Biosynthesis, (Dahlberg A. and Zimmerman B., eds.), CRC Press, Boca Raton, FL pp. 111-128), the remaining RNA electron density neatly accommodated the sequence of 5S rRNA. The interpretation of protein electron density corresponding to the protein was more complicated because each protein region had to be identified chemically before the appropriate sequence could be fit into it, but about 4,000 amino acid residues were fit into electron density.

The *H. marismortui* 50S subunit appears to contain thirty-one proteins, and there are sequences in the Swiss-Prot data bank for twenty eight of those thirty one proteins, including one, HMS6 or L7ae, that was originally assigned to the small ribosomal subunit (Whittmann-Liebold *et al.* (1990) *supra*). The three remaining proteins were identified using the sequences of the ribosomal proteins from eukaryotes and other archeal species as guides. No electron

density was found for one of the *H. marismortui* large ribosomal subunit proteins in the sequence database, LX. Either the assignment of LX to the large subunit is in error, or LX is associated with a disordered region of the subunit, or LX is absent from the subunits examined altogether.

The 2.4 Å resolution electron density map lacks clear electron density for proteins L1, L10, L11 and L12, the positions of which are known from earlier low resolution X-ray and/or electron microscopic studies. These proteins are components of the two lateral protuberances of the subunit, which are both poorly ordered in these crystals. L1 is the sole protein component of one of them (Oakes, M. *et al.* (1986)), Structure, Function and Genetics of Ribosomes, (Hardesty, B. and Kramer, G., eds.) Springer-Verlag, New York, NY, 47-67) and is evident in 9 Å resolution density maps of the subunit (Ban *et al.* (1998) *supra*), but not at higher resolutions. L10, L11 and L12 are components of the other protuberance, which is often referred to as the L7/L12 "stalk" (Oakes *et al.* (1986) *supra*). L11 and the RNA to which it binds were located in the 5 Å resolution electron density map of the *H. marismortui* large subunit (Ban *et al.* (1999) *supra*) using the independently determined crystal structures of that complex (Conn GL *et al.* (1999) *Science* 284: 1171-1174; Wimberly *et al.* (1999) *Cell* 97: 491-502). A protein fragment (about 100 residues) that is associated with the RNA stalk that supports the L11 complex can be seen in the 2.4 Å resolution map. Based on location, it must be part of L10. There is no electron density corresponding to L12 seen at any resolution, but the L12 tetramer is known to be attached to the ribosome through L10, and the L10/L12 assembly is known to be flexible under some circumstances (Moller *et al.* (1986) Structure, Function, and Genetics of Ribosomes, *supra*, pp. 309-325), which may explain its invisibility here.

The structures of eubacterial homologues of proteins L2, L4, L6, L14, and L22 have previously been determined in whole or in part (see, Table 2). L2, L6 and L14 were initially located in the 5 Å resolution map (Ban *et al.* (1999) *supra*). L4 and L22 have now been identified and positioned the same way. Electron density corresponding to most of the remaining proteins was assigned by comparing chain lengths and sequence motifs deduced from the electron density map with known sequence lengths, guided by the information available about relative protein positions (Walleczek *et al.* (1988) *EMBO J.* 7: 3571-3576) and protein interactions with 23S rRNA and 5S rRNA (Ostergaard *et al.* (1998) *J. Mol. Biol.* 284: 227-240).

Each of the protein electron density regions so identified is well accounted for by its amino acid sequence.

The most interesting of the proteins identified by sequence similarity was L7ae, which first appeared to be L30e. The L30e identification seemed plausible because the structure of yeast L30 superimposes neatly on the electron density of L7ae, and the structure of the RNA to which L7ae binds closely resembles that of the RNA to which yeast L30 binds (Mao, H. *et al.* (1999) *Nat. Struct. Biol.* 6: 1139-1147). Nevertheless, the sequence of HMS6, which by sequence similarity is a member of the L7ae protein family, fits the electron density better. Four of the other proteins identified by sequence similarity, L24e, L37e, L37ae, and L44e, contain zinc finger motifs. The rat homologues of L37e and L37ae were predicted to be zinc finger proteins on the basis of their sequences (Wool *et al.* (1995) *supra*), and this prediction helped identify their homologues in *H. marismortui*.

Table 2

Large Subunit Proteins from *Haloarcula Marismortui*

				Interactions[5]							
Name[1]	Hmlg[2]	Lgth[3]	Conf[4]	1	2	3	4	5	6	5S	Proteins
L1*		211	glb.					+			none
L2†	RL8	239	glb+ext		+	+	+	+			(L37ae)
L3	RL3	337	glb+ext		+		+	+	+		L14, L24e, (L13)
L4†	RL4	246	glb+ext	+	+			Å			(L18e), (L24), (L37e)
L5	RL11	176	glb					+		+	L18
L6	RL9	177	glb		Å			Å	+		(L13)
L10*	RP0	348	glb?		+						L12
L11*	RL12	161	glb		+						none
L12*	RL1/2	115	glb								L10
L13	RL13a	145	glb		+			Å	Å		(L3), (L6)
L14	RL23	132	glb				+	+	+		L3, L24e
L15	RL27a	164	glb+ext	+	+			+			(L18e), (L32e)
L18	RL5	186	glb+ext		Å			+		+	L5, L21e
L19	RL19	148	glb+ext		+	+	+		Å		none
L22	RL17	154	glb+ext	+	Å	+	+	+	+		none
L23	RL23a	84	glb	Å		+					L29, (L39e)
L24	RL26	119	glb+ext	+							(L4)
L29	RL35	70	glb	+							L23
L30	RL7	154	glb		+					+	none
L18e	RL18	115	glb		+						(L4), (L15)
L21e	RL21	95	glb		+			+		Å	L18
L24e	RL24	66	glb				Å		+		L3, L14
L31e	RL31	91	glb			+	+		+		none
L32e	RL32	240	glb	Å	+						(L15)
L37e	RL37	56	glb+ext	+	+	+	Å				(L4)
L39e	RL39	49	ext	+		+					(L23)
L44e	RL36a	92	glb+ext	+			Å	+			(L15e)
L7ae	R17a	110	glb	Å							L15e
L10e	RL10	163	glb		+			+		Å	none
L15e	RL15	184	glb+ext	+	Å	Å	Å	+			(L44e), L7ae
L37ae	RL37a	72	glb+ext		+	+	+				L2

The top block of proteins include all those known to have eubacterial homologues of the same name. The second block lists proteins found in the *H. marismortui* large ribosomal subunit that have only eukaryotic homologues (Wittmann-Liebold *et al.* (1990) *supra*). Their names are all followed by the letter "e" to distinguish them from eubacterial proteins that would otherwise have the same name. The third block are large subunit proteins for which no *H. marismortui* sequence yet exists. They are identified by sequence homology using standard L names.

[1] The structures of all or part of homologues of the following proteins were previously determined: L1 (Nevskaya *et al.* (2000) *Struct. Fold. Des.* 8: 363), L2 (Nakagawa, A. *et al.* (1999) *EMBO J.* 18: 1459-1467), L4 (Wahl *et al.* (2000) *EMBO J.* 19: 807-818), L6 (Golden *et al.* (1993) *EMBO J.* 12: 4901-4908), L11 (Conn *et al.* (1999) *supra*; Wimberly *et al.* (1999) *supra*; Markus *et al.* (1997) *Nature Struct. Biol.* 4: 70-77), L12 (Leijonmarck, M. *et al.* (1980) *Nature* 286: 824-827), L14 (Davies *et al.* (1996) *Structure* 4: 55-66), L22 (Unge *et al.* (1998) *Structure* 6: 1577-1586), L30 (Wilson *et al.* (1986) *Proc. Nat. Acad. Sci. USA* 83: 7251-7255). All other structures, except 10, have been newly determined in this study.

[2] Rat homologue. Rat equivalents to *H. marismortui* proteins are from (Mao *et al.* (1999) *supra*).

[3] Sequence chain length.

[4] Conformation: glb = globular; ext = extension
[5] The protein interactions with the 6 domains of 23S rRNA, 5S rRNA and other proteins are specified. (+) implies that the interaction is substantial. (Å) implies a weak, tangential interaction. Protein names are shown in parentheses implies that the interactions are weak; otherwise, the interaction is substantial.

*All entries so designated describe proteins that are not fully represented in the electron density maps described here. The summary information provided is derived from literature sources and is included here for completeness only.
† The structure available for this protein in isolation does not include the extension(s) reported here.

3. General Appearance of the Subunit.

In its crown view (see Figure 2), the large ribosomal subunit, which is about 250 Å across, presents its surface that interacts with the small subunit to the viewer with the three projections that radiate from that surface pointed up. Although the protuberance that includes L1 is not visible in the 2.4 Å resolution electron density map, the structure of L1, which has been determined independently (Nikonov *et al.* (1996) *EMBO J.* 15: 1350-1359), has been positioned approximately in lower resolution maps (Ban *et al.* (1998) *supra*) and is included here to orient the reader. It is evident that, except for its two lateral protuberances, the large ribosomal subunit is monolithic. There is no hint of a division of its structure into topologically separate domains. In addition, partly because it lacks obvious domain substructure but also because it is so large, it is impossible to comprehend looking at it as a whole. In order to convey to the reader a sense of how it is put together, the subunit must be dissected into its chemical components.

4. RNA Secondary Structure.

All the base pairs in *H. marismortui* 23S rRNA stabilized by at least two hydrogen bonds were identified using a computer program that searched the structure for hydrogen bond donors and acceptors separated by less than 3.2 Å. Bases linked by at least two such bonds were considered paired if the angle between their normals was less than 45°, and the angle between bonds and base normals was also less than 45°. Based on the results of this analysis, a secondary structure diagram has been prepared in the format standard for 23S/28S rRNAs (see Figure 3). The secondary structure predicted for this molecule by phylogenetic comparison was remarkably accurate, but it did not find all of the tertiary pairings and failed to identify interactions involving conserved bases. In addition to base pairs of nearly every type, the RNA contains numerous examples of well-known secondary structure motifs such as base triplets, tetraloops, and cross-

strand purine stacks, but no dramatically new secondary structure motifs have been identified so far.

The secondary structure of this 23S rRNA consists of a central loop that is closed by a terminal stem, from which 11 more or less complicated stem/loops radiate. It is customary to describe the molecule as consisting of 6 domains, and to number its helical stems sequentially starting from the 5' end (see Figure 4) (Leffers *et al.* (1987) *supra*). The division of the molecule into domains shown in Figure 4 deviates from standard practice with respect to helix 25, which is usually considered to be part of domain I. It is placed in domain II because it interacts more strongly with domain II than it does with the other elements of domain I.

There are five sequences longer than 10 nucleotides in 23S rRNA whose structures cannot be determined from the 2.4 Å resolution map due to disorder. Altogether they account for 207 out of the 232 nucleotides missing from the final model. The disordered regions are: (1) all of helix 1, (2) the distal end of helix 38, (3) helix 43/44 to which ribosomal protein L11 binds, (4) the loop end of stem/loop 69, and (5) helix 76/77/78, which is the RNA structure to which L1 binds. For completeness, these regions are included in Figure 3 (in gray) with their secondary structures determined phylogenetically.

5. *Overall Architecture of rRNA.*

The six domains of 23S rRNA and 5S rRNA all have complicated, convoluted shapes that nevertheless fit together to produce a compact, monolithic RNA mass (see Figure 4(A) and 4(B)). Thus despite the organization of its RNAs at the secondary structure level, in three-dimensions, the large subunit is a single, gigantic domain. In this respect, it is quite different from the small subunit, which is a flatter object that is not at all monolithic. Even in low resolution electron micrographs the small subunit consists of three structural domains, each of which, it turns out, contains one of the three secondary structure domain of its RNA (Noller *et al.* (1990) The Ribosome: Structure, Function, and Evolution, *supra*, pp. 73-92). This qualitative difference between the two subunits may reflect a requirement for conformational flexibility that is greater for the small subunit.

Domain I, which looks like a mushroom (see Figure 4(E)), lies in the back of the particle, behind and below the L1 region. The thin part of the domain starts in the vicinity of domain VI,

which is where its first and last residues are located. Helices 1 and 25 span the particle in the back and then the domain expands into a larger, more globular structure below and behind the L1 region.

The largest of the six 23S rRNA domains, domain II, which accounts for most of the back of the particle, has three protrusions that reach towards the subunit interface side of the particle (see Figure 4(F)). One of them (helix 42 - 44) is the RNA portion of the L7/L12 stalk, which is known to interact with elongation factors, is not well-ordered in these crystals. The second domain II protrusion is helix 38, which is the longest, unbranched stem in the particle. It starts in the back of the particle, bends by about 90 degrees and protrudes towards the small subunit between domains V and 5S rRNA. The third region, helix 32-35.1, points directly towards the small subunit and its terminus, the loop of stem/loop 34, interacts directly with the small ribosomal subunit (Culver *et al.* (1999) *Science* 285: 2133-2135). This loop emerges at the subunit interface between domains III and IV.

Domain III is a compact globular domain that occupies the bottom left region of the subunit in the crown view (see Figure 4(G)). It looks like a four pointed star with the origin of the domain (stem/loop 48) and stem/loops 52, 57, and 58 forming the points. The most extensive contacts of domain III are with domain II, but it also interacts with domains I, IV and VI. Unlike all the other domains, domain III hardly interacts with domain V at all; the sole contact is a single van der Waals contact involving a single base from each domain.

Domain IV accounts for most of the interface surface of the 50S subunit that contacts the 30S subunit (see Figure 4(H)). It forms a large diagonal patch of flat surface on that side of the subunit, and connects to domains III and V in the back of the particle. Helices 67-71 are the most prominent feature of domain IV, and form the front rim of the active site cleft, which is clearly visible at low resolution (see Figure 2). This is one of the few regions of the 23S rRNA that is not extensively stabilized by ribosomal proteins. Helix 69 in the middle of this ridge interacts with the long penultimate stem of 16S rRNA in the small ribosomal subunit and can be viewed as a divider separating A-site bound tRNAs from P-site bound tRNAs.

Domain V, which is sandwiched between domains IV and II in the middle of the subunit, is known to be intimately involved in the peptidyl transferase activity of the ribosome. Structurally the domain can be divided into three regions (see Figures 4(I) and 4(J)). The first

starts with helix 75 and ultimately forms the binding site for protein L1. The second, which consists of helices 80-88, forms the bulk of the central protuberance region, and is supported in the back by the 5S rRNA and domain II. The third region, which includes helices 89-93, extends towards domain VI and helps stabilize the elongation factor binding region of the ribosome.

The smallest domain in 23S rRNA, domain VI, which forms a large part of the surface of the subunit immediately below the L7/L12 stalk, resembles a letter X with a horizontal bar at the bottom (see Figure 4(K)). An interesting region of this domain is the sarcin-ricin loop (SRL) (stem/loop 95), the structure of which has been extensively studied in isolation (Szewczak *et al.* (1995) *J. Mol. Biol.* 247: 81-98). The SRL is essential for factor binding, and ribosomes can be inactivated by the cleavage of single covalent bonds in this loop (Wool *et al.* (1992) *TIBS* 17: 266-269). As suggested by nucleotide protection data, the major groove of this loop is exposed to solvent (Moazed *et al.* (1988) *Nature* 334: 362-364), and its conformation is stabilized by proteins and through interaction with domain V that involves two bases on the minor grove side. The nucleotides involved are A 2699 and G 2700 in domain VI, and A 2566 and G 2567 in domain V.

5S ribosomal RNA, which is effectively the seventh RNA domain in the subunit, consists of three stems radiating out from a common junction called loop A (see Figure 4(D)). In contrast to what is seen in the crystal structure of fragment 1 from *E. coli* 5S rRNA (Correll *et al.* (1997) *Cell* 91: 705-712), the helix 2/3 arm of the molecule stacks on its helix 4/5 arm, not helix 1 (see Figure 4(L)). This arrangement results from a contorted conformation of loop A residues that involves two stacked base triples. Indeed, from the secondary structure point of view, the loopA-helix 2,3 arm of 5S rRNA is quite remarkable. Nowhere else in the subunit is there a higher concentration of unusual pairings and of convoluted RNA secondary structure.

6. Sequence Conservation and Interactions in 23S rRNA.

While 23S/28S rRNAs contain many conserved sequences, they also vary substantially in chain length. Shorter 23S/28S rRNAs are distinguished from their longer homologues by the truncation of, or even the elimination of entire stem/loops, and by comparing sequences, one can identify a minimal structure that is shared by all (Gerbi (1995) Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis, *supra*, pp. 77-88). The expansion sequences in the 23S rRNA of *H. marismortui*, *i.e.*, the sequences it contains that are larger than

the minimum, are shown in Figure 5 in green. They are largely absent from the subunit interface surface of the particle, but they are abundant on its back surface, far from its active sites. This is consistent with low resolution electron microscopic observations suggesting that the region of the large subunit whose structure is most conserved is the surface that interacts with the small subunit (Dube *et al.* (1998) *Structure* 6: 389-399).

There are two classes of conserved sequences in 23S rRNA. One contains residues concentrated in the active site regions of the large subunit. The second class consists of much shorter sequences scattered throughout the particle (Figure 5: red sequences). The SRL sequence in domain VI and the cluster of conserved residues belonging to domain V that are located at the bottom of the peptidyl transferase cleft are members of the first class. They are conserved because they are essential for substrate binding, factor binding and catalytic activity. Most of the residues in the second class of conserved residues are involved in the inter- and intra-domain interactions that stabilize the tertiary structure of 23S rRNA. Adenosines are disproportionately represented in this class. The predominance of A's among the conserved residues in rRNAs has been pointed out in the past (Ware *et al.* (1983) *Nucl. Acids. Res.* 22: 7795-7817).

In addition to its reliance on A-dependent motifs, the tertiary structure of the domains of 23S rRNA and their relative positions are stabilized by familiar tertiary structure elements like RNA zippers and tetraloop/tetraloop receptor motifs (Moore, P.B. (1999) *Annu. Rev. Biochem.* 68: 287-300), and in many places, base pairs and triples stabilize the interactions of sequences belonging to different components of the secondary structure of 23S rRNA.

Interestingly, 5S rRNA and 23S rRNA do not interact extensively with each other. The few RNA/RNA interactions there are involve the backbones of the helix 4/5 arm of 5S rRNA and the backbone of helix 38 of 23S rRNA. Most of the free energy and all of the specificity of 5S rRNA binding to the large ribosomal subunit appears to depend on its extensive interactions with proteins that act as modeling clay sticking it to the rest of ribosome.

7. Proteins in the 50S Ribosomal Subunit.

The structures of twenty seven proteins found in the large ribosomal subunit of *H. marismortui* (Table 2) have been determined. Twenty-one of these protein structures have not been previously established for any homologues, and the structures of the six that do have

homologues of known structure have been rebuilt into the electron density map using their *H. marismortui* sequences. In addition, there are structures available for homologues of *H. marismortui* L1, L11 and L12, which cannot be visualized in the 2.4 Å resolution electron density map. Only the structure of L10 is still unknown among the thirty one proteins known to be present.

Not every one of these structures is complete. For example, an entire domain of L5 is missing from the electron density, presumably because of disorder. L32e is also noteworthy in this regard. About twenty residues from its N-terminus are not seen in the electron density map, and the electron density map suggests that its C-terminal residue is covalently bound to the most N-terminal of its visible residues.

Of the thirty large subunit ribosomal proteins whose structures are known, 17 are globular proteins, similar in character to thousands whose structures are in the Protein Data Bank (Table 2). The remaining thirteen proteins either have globular bodies with extensions protruding from them ("glb+ext") or are entirely extended ("ext"). Their extensions often lack obvious tertiary structure and in many regions are devoid of secondary structure as well (see Figure 6). These extensions may explain why many ribosomal proteins have resisted crystallization in isolation. The exceptions that prove the rule are L2 and L4, both of which are proteins belonging to the "glb+ext" class. Protein L2 was crystallized and its structure solved only after its extensions had been removed (Nakagawa *et al.* (1999) *supra*), and one of the two regions of L4 that are extended in the ribosome is disordered in the crystal structure of intact L4 (Wahl *et al.* (2000) *supra*).

Except for proteins L1, L7, L10 and L11, which form the tips of the two lateral protuberances, the proteins of the 50S subunit do not extend significantly beyond the envelope defined by the RNA (see Figure 7). Their globular domains are found largely on the particle's exterior, often nestled in the gaps and crevices formed by the folding of the RNA. Thus, unlike the proteins in spherical viruses, the proteins of the large ribosomal subunit do not form a shell around the nucleic acid with which they associate, and unlike the proteins in nucleosomes, they do not become surrounded by nucleic acid either. Instead, the proteins act like mortar filling the gaps and cracks between "RNA bricks".

The distribution of proteins on the subunit surface is nearly uniform, except for the active site cleft and the flat surface that interacts with the 30S subunit. In the crown view the proteins lie around at the periphery of the subunit (see Figure 7(A)), but when viewed from the side opposite the 30S subunit binding site (the "back side"), they appear to form an almost uniform lattice over its entire surface (see Figure 7(B)). Similarly, the bottom surface of the subunit, which includes the exit of polypeptide tunnel, is studded with proteins (see Figure 7(C)). Indeed, the 5 proteins that surround the tunnel exit may play a role in protein secretion since they are part of the surface that faces the membrane and the translocon when membrane and secreted proteins are being synthesized.

Although Figure 7 shows protein chains disappearing into the ribosome interior, the degree to which proteins penetrate the body of the particle can only be fully appreciated when the RNA is stripped away. The interior of the particle is not protein-free, but it is protein-poor compared to the surface of the particle. Extended tentacles of polypeptide, many of which emanate from globular domains on the surface, penetrate into the interior, filling the gaps between neighboring elements of RNA secondary structure (see Figure 8(E)). The bizarre structures of these extensions are explained by their interactions with RNA.

Although extended, non-globular structures are rare in the protein data base, they are not unknown. Extended protein termini often form inter-protein contacts, *e.g.,* in viral capsids, presumably adopting fixed structures only upon capsid formation. The basic "tails" of histones may behave the same way when nucleosomes form. The N-terminal sequences of capsid proteins are often positively charged, and in virus crystal structures, the electron density for these sequences often disappears into the interior of the virus where these sequences presumably interact with asymmetrically arranged nucleic acid. The interactions observed in the ribosome could be useful models for these viral interactions.

The interactions of extended polypeptides and RNA in the large subunit, which stabilizes its massive nucleic acid structure, result in a tangle of RNA and protein in the center of the subunit (see Figures 8(A) and 8(B)). It is hard to imagine such an object assembling from its components efficiently in anything other than a highly ordered manner. Chaperones may well be required to prevent the aggregation of the extended regions of these proteins, which are likely to be disordered outside the context provided by rRNA, and to manage the folding of rRNA.

8. Protein and RNA Interactions.

Because protein permeates the large subunit to a surprising degree, there are only a few segments of the 23S rRNA that do not interact with protein at all. Of the 2923 nucleotides in 23S rRNA, 1157 nucleotides make at least van der Waals contact with protein (see Figure 8(D)), and there are only ten sequences longer than twenty nucleotides in which no nucleotide contacts protein. The longest such sequence contains forty-seven nucleotides, and is the part of domain IV that forms the ridge of the active site cleft.

The extent of the interactions between RNA and protein that occur when the large subunit assembles can estimated quantitatively. Using the Richards algorithm (Lee, B. *et al.* (1971) *J. Mol. Biol.* 55: 379-400) and a 1.7 Å radius probe to compute accessible surface areas, it can be shown that 180,000 $Å^2$ of surface becomes buried when the subunit forms from its isolated, but fully structured components. This is about half their total surface area. The average is about 6,000 $Å^2$ per protein. While this is an enormous amount compared to the surface buried when most protein oligomers form, it should be recognized that ribosome assembly must be accompanied by a large loss in conformational entropy that does not occur when most proteins oligomerize. The extended protein termini and loops of the ribosomal proteins are almost certainly flexible in isolation, and in the absence of protein, the RNA is probably quite flexible as well. Thus, the burial of a large amount of surface area may be required to provide the energy necessary to fix of the structures of these molecules.

All of the proteins in the particle except L12, interact directly with RNA and all but seven of the remaining thirty proteins interact with two rRNA domains or more (Table 2). The "champion" in this regard is L22, which is the only protein that interacts with RNA sequences belonging to all 6 domains of the 23S rRNA (see Figure 8(C)). The protein-mediated interactions between 5S rRNA and 23S rRNA are particularly extensive. Protein L18 attaches helix 1 and helix 2/3 of 5S rRNA to helix 87 of 23S rRNA. Protein L31e mediates an interaction between the same part of 5S rRNA and domains II and V. Loop C is linked to domain V by protein L5 and loop D is attached to domains II and V by protein L10e. Whatever else they may do, it is evident that an important function of these proteins is stabilization of the relative orientations of adjacent RNA domains. Several also help secure the tertiary structures of the domains with which they interact.

Since most ribosomal proteins interact with many RNA sequences and the number of proteins greatly exceeds the number of RNA domains, it can hardly come as a surprise that every rRNA domain interacts with multiple proteins (Table 2). Domain V, for example, interacts with thirteen proteins, some intimately and a few in passing.

It is clear that the oligonucleotide binding experiments long relied on for information about the RNA binding properties of ribosomal proteins have underestimated their potential for interacting with RNA. The high-affinity RNA binding site identified on a protein by such an experiment may indeed be important for ribosome assembly, but its many, weaker interactions with other sequences are likely to be missed, and they too may be vital for ribosome structure. Most ribosomal proteins crosslink RNA and crosslinking is impossible without multiple interactions. Similar considerations may apply to proteins that are components of other ribonucleoproteins such as the sliceosome.

Of the seven proteins that interact with only one domain, three (L1, L10, L11) participate directly in the protein synthesis process. Rather than being included in the ribosome to ensure that the RNA adopts the proper conformation, it seems more appropriate to view the RNA as being structured to ensure their correct placement. Another three (L24, L29, L18e) interact with several secondary structure elements within the domains to which they bind, and presumably function to stabilize the tertiary structures of their domains. The last of the single RNA domain proteins, L7ae, is puzzling. On the one hand, it cannot function as an RNA stabilizing protein because it interacts with only a single, short sequence in domain I, and on the other hand, it is far from the important functional sites in the subunit, the peptidyl transferase site and factor binding site. It is quite close to L1, however, which appears to be important for E-site function (Agrawal *et al.* (1999) *J. Biol. Chem.* 274: 8723-8729), and maybe it is involved in that activity.

While many ribosomal proteins interact primarily with RNA, a few interact significantly with other proteins. The most striking structure generated by protein-protein interactions is the protein cluster composed of L3, L6, L14, L19 and L24e that is found close to the factor binding site. The protein surface they provide may be important for factor interactions.

The structure presented above illuminates both the strengths and weaknesses of approaches to complex assemblies that depend on determining the structures of components in isolation. The structures of the globular domains of homologues of the proteins in large

ribosomal subunit from *H. marismortui* are largely the same as those of the corresponding domains in the intact subunit, though adjustments in domain positions are sometimes required. Consequently, these structures were very useful for locating proteins and interpreting lower resolution electron density maps. However, for obvious reasons, the structures of the extended tails and loops of ribosomal proteins cannot be determined in the absence of the RNAs that give them structure, and the feasibility of strategies that depend on producing low molecular weight RNA-protein complexes that have all the RNA contacts required to fix the structures of such proteins seems remote. RNA is also a problem. While the sarcin/ricin loop has much the same structure in isolation as it does in the ribosome, the structure of 5S rRNA in isolation differs in some respects from what is seen in the ribosome, and the structure of the isolated P-loop (Puglisi *et al.* (1997) *Nat. Struct. Biol.* 4: 775-778) does not resemble the structure of the P-loop in the ribosome at all. Clearly a "structural genomics" approach to the ribosome, which would have entailed determining the structures of all its proteins and all possible rRNA fragments, would not have worked. It may not be successful for other macromolecular assemblies either.

B. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis

Analysis of the atomic co-ordinates discussed in section IIA above together with additional atomic co-ordinates of a ribosomal subunit complexed with various analogues, similarly refined, permit an analysis of ribosome function. Accordingly, the present invention is also based on the crystals of *Haloarcula marismortui* 50S ribosomal subunit complexed either with the Yarus transition state analogue, CCdA-p-Puro, or with a mini-helix analogue of an aminoacyl-tRNA. The present invention provides the structures of both complexes. The atomic co-ordinates of the structure of both complexes were deposited on July 26, 2000, at Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Berman *et al.* (2000) *Nucleic Acid Research* 28: 235-242; http:// www.rcsb.org/pdb/) with accession numbers PDB ID: 1FFZ (50S ribosome/ CCdA-p-Puro complex) and PDB ID: 1FG0 (50S ribosome/aa-tRNA analogue).

As discussed below, the complete atomic structures of the large ribosomal subunit and its complexes with two substrate analogues show that the ribosome is a ribozyme. The complete atomic structures also provide information regarding the catalytic properties of its all-RNA active site. Both substrate analogues are contacted exclusively by conserved rRNA residues

from domain V of 23S rRNA; there are no protein side-chains closer than about 18 Å to the peptide bond being synthesized. The mechanism of peptide bond synthesis appears to resemble the reverse of the deacylation step in serine proteases, with the base of A2486 (A2451) in *E. coli* playing the same general base role as His57 in chymotrypsin. The unusual pKa that A2486 must possess to perform this function probably derives from its hydrogen bonding to G2482 (G2447) which interacts with a buried phosphate that could stabilize the unusual tautomers of two bases which is required for catalysis. The polypeptide exit tunnel is largely formed by RNA but has significant contributions from proteins L22, L39 and L4 and its exit is encircled by proteins L19, L22, L23a, L24 and L29.

The CCdA from the Yarus analogue binds to the so-called P-loop and hence must be in the P-site. Only the terminal-CCA of the aa-tRNA analogue is visible, but since it interacts appropriately with the A-loop (Kim *et al.* (1999) *Molec. Cell* 4: 859-864), it must be in the A-site. The puromycin group occupies the same location in both structures, and there are no proteins near that site. Hence, the catalytic activity of the active site must depend entirely on RNA. The N3 of A2486 (*E. coli* A2451) is the titratable group nearest to the peptide bond being synthesized and is likely functioning as a general base to facilitate the nucleophilic attack by the α-amino group of the A-site substrate. In order to function in this capacity, the pKa of this base has to be roughly 5 units higher than normal.

1. Structures of Substrate Analogue Complexes.

In order to establish how substrates interact at the A-site and P-site of the large subunit, two substrate analogues were used. One of the analogues, which was designed to mimic the acceptor stem of an aa-tRNA and bind to the A-site, was a twelve base-pair RNA hairpin with an aminoacylated, four-nucleotide extension on its 3' end (see Figure 9). The sequence used was that of the tRNA tyr acceptor stem and it is terminated with puromycin, which itself is an analogue of tyrosyl-A76. The second analogue used was the Yarus transition state analogue, CCdA-p-puromycin. As in the case of the A-site substrate analogue, the puromycin of the Yarus inhibitor is expected to bind at the A-site, while its CCdA moiety should bind at the P-site.

The positions of the Yarus inhibitor and the tRNA acceptor stem analogue were determined by soaking these molecules into crystals of the *H. marismortui* 50S ribosomal subunit, measuring diffraction data to 3.2 Å resolution and calculating difference electron density

maps (Welch *et al.* (1997) *Biochemistry* 36: 6614-6623). Maps of the complexes were also calculated using $2F_o$(complexed)- F_o(uncomplexed) as coefficients, to examine the shifts in the positions of ribosome residues that occur when these analogues bind (see Figure 10(B) and Table 3).

Table 3.
Statistics for Data Collection and Scaling.

Crystal	Native A	Native B	CcdAp-Puro	Mini-helix
Soak time (hours)	-	-	24	24
Soak concentration (μM)	-	-	100	100
Wavelength (Å)	1.0	1.0	1.0	1.0
Observations	1,571,171	1,344,877	2,590,726	2,712,813
Unique	284,533	369,167	367,284	447,121
Redundancy	5.5	3.6	7.0	6.0
Resolution limits (Å)	70.0-3.2	70.0-3.0	70.0-3.0	70.0-2.8
(High-resolution bin)*	(3.26-3.20)	(3.05-3.00)	(3.23-3.17)	(3.08-3.02)
Completeness	94.1 (96.0)	98.9 (99.3)	98.6 (99.9)	99.6 (100)
$I/\sigma I$	14.6 (4.0)	10.8 (3.1)	11.0 (2.8)	10.7 (2.9)
R_{merge}†	10.2 (40)	11.5 (38)	18.8 (84)	14.3 (72)
R_{iso} Native A‡	-	-	6.8 (20.8)	14.4 (25.2)
R_{iso} Native B‡	-	-	12.6 (27.4)	17.5 (31.0)

*Statistics in parenthesis are calculated for the high-resolution bin used in map calculations, which, as indicated was sometimes lower in resolution than the high-resolution bin used in data reduction. † R_{merge}: $\Sigma\Sigma_i|I_{(h)} - I_{(h)i}|I\Sigma\Sigma I_{(h)i}$' where $I_{(h)}$ is the mean intensity after reflection. ‡R_{iso}: $\Sigma|F_{PH} - F_P|\Sigma F_{PH}$_ where F_{PH} and F_P are the soaked and the native crystal structure factor amplitudes respectively.

A model for the entire Yarus inhibitor could be fitted into the difference density (see Figure 10(A)), and the electron density map of the complex shows the N3 of A2486 (2451) within hydrogen bonding distance of a non-bridging oxygen of the phosphoramide (see Figure 10(B)). The inhibitor's two C's, which correspond to C74 and C75 of peptidyl-tRNA, are Watson-Crick base-paired with G2285 (2252) and G2284 (2251)in the P-loop, respectively (see Figure 11(A)). The C74-G2285 (2252) interaction was predicted by the results of Noller and coworkers (Noller *et al.* (1992) *Science* 256: 1416-1419). The dA, which corresponds to A76 of a tRNA in the P-site, is not base-paired, but rather stacks on the ribose of A2486 and hydrogen bonds to the 2'OH of nucleotide A2485 (see Figure 12(B)).

Only the CC-puromycin moiety of the mini-helix acceptor stem analogue showed ordered electron density in its difference electron density map (see Figure 10(C)). The C75 of the acceptor stem CCA is Watson-Crick base-paired with G2588 (2553) of the A-loop, whereas the C74 is more disordered and not base-paired but appears to stack on a ribosome base. The dimethyl A of the A-site inhibitor puromycin is positioned identically to the dimethyl A of the

Yarus inhibitor. Further, the dimethyl A of puromycin, which is the A76 equivalent of an A-site tRNA, interacts with the A-loop in much the same way that the A76 from the P-site CCA interacts with the P-loop (see Figure 11(B)).

The most notable of the several conformational changes in the ribosome induced by the binding of the transition state analogue is the ordering of base A2637 (2602), which is disordered in the unliganded enzyme (see Figure 11(B)). It becomes positioned between the CCA bound at the A-site and the CCA bound at the P-site. The base of U2620 (2585) also moves so that it can make a hydrogen bond with the 2' hydroxyl of the ribose of A76 in the A-site, and U2619 and G2618 shift to allow the A76 to be positioned. Smaller shifts in positions are observed in the positions of A2486, whose N3 is near to the non-bridging oxygen of the phosphate, and one of the G residues with which it interacts, G2102 (2482).

2. Location and Chemical Composition of the Peptidyl Transferase Site.

The inhibitors are bound to a site made entirely of 23S rRNA with no proteins nearby, proving that the ribosome is a ribozyme. Both the Yarus inhibitor and the A-site analogue of aa-tRNA bind to the large subunit at the bottom of a large and deep cleft at the entrance to the 100 Å long polypeptide exit tunnel that runs through to the back of the subunit (see Figure 12). This site is surrounded by nucleotides belonging to the central loop of 23S RNA domain V, the "peptidyl transferase loop." Nucleotides from the single stranded portions of this loop make the closest approach to the phosphate that mimics the tetrahedral carbon intermediate. In general, the helices that extend from the peptidyl transferase loop in 2 secondary structure diagrams of 23S rRNA also extend away from the active site in the tertiary structure (see Figure 13). Although there are 13 proteins that interact with domain V (see Figure 14(A)), there are no globular proteins in the vicinity of the inhibitor. The closest polypeptides are the non-globular extensions of several proteins (L2, L3, L4, L10e) that penetrate deeply into domain V and approach the active site (see Figure 14(B)). These extensions fill many of the voids between the RNA helices of domain V, neutralize phosphate backbone charge, and presumably stabilize the structure of the domain and its association with other RNA regions. However, none of their side chain atoms is closer than about 18 Å to the phosphorus of the inhibitor's phosphate group, which marks the site where peptide bonds form. Furthermore, the substrate analogue is completely enclosed in an rRNA cavity that is so tightly packed that there is no possibility that

an unidentified peptide could be lurking nearby (see Figure 15). Thus, the catalytic entity in the ribosome must be RNA.

Two of the proteins with long termini or loops penetrating the rRNA scaffold of domain V are proteins that could not previously be excluded from involvement in the peptidyl transferase reaction L2 and L3 (Noller (1991) *Ann. Rev. Biochem.* 60: 191-227). Noller and colleagues (Noller *et al.* (1992) *supra*) found that under conditions which prevent RNA denaturation, extensive digestion of *Thermus thermophilus* 50S subunits with proteases followed by extraction with phenol and other agents that disrupt protein-RNA interactions did not remove several peptides from the subunit that were less than 10,000 in molecular weight. The structure makes it clear why these protein fragments were particularly resistant to protease treatments. While protease treatment could digest the globular protein domains on the surface of the large subunit, it could not remove the long termini or loops that penetrate deeply in the 23S rRNA because they are sequestered within the rRNA and thus protected from cleavage, independently of the globular domains.

3. Peptidyl Transferase Active Site.

The RNA that surrounds the substrate analogues is closely packed, much like the active site region of a protein enzyme and the nucleotides in contact with the inhibitor are greater than 95% conserved in all three kingdoms of life (see Figure 15). Thus, it is clear that the ribosome is a ribozyme, but what gives the RNA its catalytic power?

Without wishing to be bound by theory, the residue most likely to be involved in catalysis, presumably as a general base, is A2486, whose N3 is about 3 Å from the phosphoramide oxygen of the Yarus inhibitor that is the analogue of the carbonyl oxygen of a nascent peptide bond and about 4 Å from the amide that corresponds to the amide nitrogen of the peptide bond being synthesized. Ordinarily, the pKa of the N1 of adenosine monophosphate is about 3.5 and that of its N3 is perhaps 2 pH units lower (Saenger (1984) *Principles of Nucleic Acid Structure*, (C.R. Cantor, eds.), Springer Advanced Texts in Chemistry, Springer-Verlage, New York, NY), and in order for A2486 to function as a general base, its pKa would have to be raised to 7 or higher. The crystal structure itself suggests that its pKa is, in fact, quite unusual. The N3 of A2486 can only hydrogen bond to the phosphate oxygen, as observed, if it (or the phosphate oxygen) is protonated. The distance between these two atoms is about 3 Å indicating

that a hydrogen bond does, indeed, exist between them. Since the crystal is at pH 5.8, this implies that the pKa of the N3 is greater than 6. Muth and Strobel have measured the pKa of the corresponding A in *E. coli* 23S RNA by examining its dimethyl sulfate reactivity as a function of pH and have concluded that it is 7.6, though they cannot be sure from their experiments whether it is the N3 or N1 whose pKa they have measured (Muth *et al.* (2000) *Science* 289: 947-950). Because there is no other available, titratable RNA functional group closer than about 7 Å to the nascent peptide bond, there is not other group available to function as a general base.

There are several features of environment of A2486 (2451) that might affect its pKa. The pKa of the N3 of A2486 (2451) may be increased significantly in part by a charge relay mechanism, analogous to that which occurs in the active site of the serine proteases, with the buried phosphate of A2485 (2450) performing a similar function as the buried carboxylate of Asp102 of chymotrypsin. The experimental 2.4 Å electron density map unambiguously establishes the hydrogen bonding interactions in this most critical region of the active site (see Figure 16(A)). The N6 of A2486 interacts with the O6 atoms of G2482 (2447) and G2102 (2061) (see Figure 16(B)). The N2 of G2482 is also interacting with a non-bridging oxygen of the phosphate group of A2485 (2450) that is among the total of 3 most solvent inaccessible phosphate groups (826, 1497 and 2485) in the large ribosomal subunit for which we do not see any neutralizing counterion in the 2.4 Å resolution map. Weak density that may correspond to a water molecule is hydrogen bonded to the other non-bridging oxygen. A neutralizing counterion is not apparent in this structure. The buried phosphate of A2485 could abstract the proton from the exocyclic N2 of G2482 in order to neutralize its energetically unfavorable buried negative charge. This, in turn, would stabilize the otherwise rare imino tautomer of that base. The interaction of an imino of G2482 with A2486 likewise can stabilize the imino tautomer of A2486 that would result in a negative charge on its N3 were it unprotonated (see Figure 16(C)). In this way, some of the negative electrostatic charge originating on the buried phosphate of A2485 could be relayed to the N3 of A2486, thereby increasing its pKa.

A second feature of the environment of the catalytic site that may affect its stability, tautomeric state and electrostatic charge distribution is a bound monovalent cation. A potassium or a sodium ion interacts with the O6 of G2482 and G2102 as well as with three other bases. Its identity as a potassium ion is established by its observed continuation and by an independent

experiment showing that a rubidium ion can bind to this site. The monovalent ion might also stabilize non-standard tautomers, but its expected influence on the pKa of A2486 is less obvious. Early biochemical experiments have shown the importance of potassium for peptidyl transferase activity (Monro (1967) *supra*; Maden *et al.* (1968) *supra*) and this binding site could be responsible for that affect.

It may also be the case that stabilization of an imino tautomer by a buried phosphate explains the expected higher pKa of a catalytic cytosine in the active site of the hepatitis delta virus ribozyme (Ferre-D'Amare *et al.* (1998) *Nature* 395: 567-574; Naharo *et al.* (2000) *Science* 287: 1493-1497). In this case, a backbone phosphate, whose solvent accessibility is similar to that of A2485 in the ribosome, is observed to hydrogen bond to the N4 of C, and the protonated form of the imino tautomer of that C would neutralize the phosphate, promoting the function of its N3 as a general acid (Naharo *et al.* (2000) *supra*).

4. Catalytic Mechanism of Peptide Bond Formation.

The proximity of the N3 of A2486 (2451) to the peptide bond being synthesized and the nature of the reaction catalyzed suggest a chemical mechanism of peptide synthesis that is analogous to the reverse of the deacylation step seen in serine proteases during peptide hydrolysis (Blow *et al.* (1969) *Nature* 221: 337-340; Steitz *et al.* (1982) *Ann. Rev. Biophys. Bioeng.* 11: 419-444). In that mechanism, the basic form of His57 abstracts a proton from the α-amino group of the peptide product as it attacks the acyl-Ser195. Formation of the tetrahedral carbonyl carbon intermediate is stabilized by interaction of the oxyanion formed with backbone amides (the "oxyanion hole"); His57 shuttles the proton acquired from the α-NH_2 to Ser195 as the tetrahedral intermediate breaks down.

The residue A2486 (2451) appears to be the analogous to His57 in chymotrypsin and that the peptidyl-tRNA is analogous to acyl-Ser195. Thus, the N3 of A2486, with its greatly elevated pKa, abstracts a proton from the α-amino group of the A-site bound aminoacyl-tRNA facilitating the nucleophilic attack of that amino group on the carbonyl carbon that acylates the 3' OH of the tRNA in the P-site (see Figure 17(A)). In contrast to the serine proteases, however, the oxyanion of the tetrahedral intermediate is near to the protonated N3 of A2486 (A2451) rather than being proximal to a separate oxyanion binding site. Thus, it could be that the protonated N3 of A2486 stabilizes the formation of the oxyanion by hydrogen bonding to it, as we observe in the Yarus

inhibitor complex (see Figure 17(B)). The N3 of A2486 could then subsequently transfer its proton to the 3' hydroxyl of the P-site bound tRNA, which is liberated as the peptide shifts to the A-site bound tRNA (see Figure 17(C)).

An additional question is how is the catalyzed hydrolysis of the peptidyl tRNA in the P-site prevented prior to the delivery of the next appropriate aa-tRNA to the A-site? It appears from this complex that water would not be excluded from access to the peptidyl link to the P-site tRNA if the A-site were vacant. An analogous problem was discussed by Koshland in the 1960s (Koshland, Jr. (1963) *Cold Spring Harbor Symp. Quant. Biol.* 28: 473-489), who theorized why hexokinase does not hydrolyze ATP in the absence of glucose, since water should bind perfectly well to the binding site used by the 6-hydroxyl of glucose. The answer proposed was induced fit, *i.e.,* hexokinase is not catalytically competent until the glucose binds and produces a conformational change that orients substrates and catalytic groups optimally. This indeed appears to be the case (Bennett, Jr. *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75: 4848-4852). Similarly, it could be that the catalytic A2486 and/or the peptidyl substrate are not properly oriented or that the binding site for the α-NH_2 group is blocked by a reoriented ribosome base in the absence of aa-tRNA in the A-site. We do observe that the base of U2620 is close to A2486 in the ligand free structure, and it may serve as the appropriate plug that prevents spontaneous hydrolysis of peptidyl-tRNA.

Thus, it appears that this RNA enzyme uses the same principles of catalysis as a protein enzyme. First, a large catalytic enhancement is achieved by precisely orienting the two reactants, the αNH_2 from the A-site aminoacyl-tRNA and the carbonyl carbon from the P-site peptidyl-tRNA. This may be accomplished, in part, by the interactions of the CCA ends of the A-site and P-site tRNAs with the A-loop and P-loop, respectively. Secondly, acid-base catalysis and transition state stabilization are achieved by an enzyme functional group (A2486 (2451) in this case) whose chemical properties are altered appropriately by the active site environment. Third, similar chemical principles may be used by RNA and protein enzymes to alter the pKa's of functional groups. A buried carboxylate of Asp102 acting through His57 alters the nucleophilicity of Ser195 in chymotrypsin (Blow *et al.* (1969) *supra*). In the ribosome a solvent inaccessible phosphate may act through G2482 (2447) alters the nucleophilicity of the N3 of

A2486 (2451). It could be that RNA molecules "learned" how to use the chemical principles of catalysis significantly before protein molecules did.

5. tRNA Binding.

While it is not possible to bind tRNA molecules to either the A-or P-sites in these crystals for steric reasons, it is possible to place the A-, P- and E-site tRNA molecules on the large ribosomal subunit in the same relative orientation that Cate *et al.* observed in their crystallographic study of the *Thermus aquaticus* 70S ribosome. The co-ordinates of the three tRNA molecules in the relative positions seen in the 70S ribosome can be docked on the *Haloarcula marismortui* large ribosomal subunit in a way that avoids steric clashes and places the acceptor stems of the A-site and P-site tRNAs near to the positions of the CCAs we have observed bound to the A-loop and P-loop (see Figure 18). Although nucleotides C74 and C75 were modeled in a different conformation in the 7.8 Å ribosome map, the C74 residues from the CCAs in both the A- and P-sites can be connected to residue 72 of the docked A-site and P-site tRNAs through a modeled residue 73, and it appears that the tRNA molecules fit well onto the surface of the subunit. Unexpectedly, this modeling places the E-site, P-site and A-site bound tRNA molecules in close proximity to three ribosomal proteins. Proteins L5 and L10e are near tRNAs in the P-site and A-site. Since both of these proteins also interact with 5S rRNA, this observation raises the possibility that 5S rRNA and some of its associated proteins might help stabilize the positioning of ribosome bound tRNAs and is consistent with the fact that 5S rRNA enhances ribosomal activity, but is not absolutely essential for it (Moore, Ribosomal RNA & Structure, Evolution, Processing and Function in Protein Biosynthesis (1996), *supra*, pp. 199-236). Protein L44e appears to interact with the E-site tRNA and may contribute to E-site activity. According to this docking experiment the A-site tRNA interacts with the highly conserved stem-loop 2502-2518 (2467-2483) which together with L10e forms a large concave surface that contacts the tRNA on the T-stem, utilizing the exact same binding site exploited by EF-Tu (Gutell *et al.* (2000) *supra*).

Examination of the relationships between the CCAs bound in the A- and P-sites and the tRNAs to which they are connected as well as their interactions with the ribosome also leads to some insights into translocation. Immediately after formation of the new peptide bond and deacylation of the P-site tRNA, the acceptor end of the P-site tRNA is known to move to the E-

side and that of the A-site tRNA moves to the P-site (Blobel *et al.* (1970) *J. Cell. Biol.* 45: 130-145). The approximate modeling of the 3 tRNA molecules on the large subunit suggests some possible contributions to this process. First, there are two base-pairs between the P-site tRNA and the P-loop and only one between the A-site and the A-loop. Moving from the A- to the P-site increases base-pairing, though there must be a concomitant attraction of the deacylated P-site tRNA to an E-site. Further, the CCAs bound to the A and P loops are related by 180° rotation, whereas the tRNAs to which they are attached are not. Thus, the relationships of the CCAs to the acceptor stems cannot be the same in both sites and may not be equally stable. If the conformation of the A-site tRNA is less stable, then moving a tRNA from the A- to the P-site might be energetically favored.

6. Polypeptide Exit Tunnel.

It appears very likely from the structure that all nascent polypeptides pass through the polypeptide exit tunnel before emerging from the ribosome, because there appears to be no other way out. We are now able to address two important questions about the functioning of the polypeptide exit tunnel: (1) Why do nascent proteins not stick to its walls? Teflon has the marvelous property of not sticking to denatured egg proteins, so how has the ribosome achieved a similar non-stick surface for the denatured proteins that must pass through the tunnel? (2) Do proteins fold to any degree in the tunnel giving the ribosome a chaperon-like function?

The length of the tunnel from the site of peptide synthesis to its exit is about 100 Å, broadly consistent with the length of nascent polypeptide that is protected from proteolytic cleavage by the ribosome (Moazed *et al.* (1989) *Nature* 342:142) and the minimum length required for antibody recognition at the exit (Picking *et al.* (1992) *Biochemistry* 31: 2368-2375). The tunnel is largely straight, except for a bend 20 to 35 Å from the peptidyl transferase center (see Figure 19). Its diameter varies from about 20 Å at its widest to a narrow point of about 10 Å at the very beginning and at a position 28 Å from the tunnel exit with an average diameter of about 15 Å. Since the smallest orifice through which the polypeptide product must pass only barely accommodates the diameter of an α-helix diameter, it seems unlikely that significant protein folding beyond the formation of α-helix could occur within the ribosome.

The majority of the tunnel surface is formed by domains I - V of 23S rRNA, but significant contributions are also made by the non-globular regions of proteins L22, L4 and L39

which not only fill some of the voids in the RNA scaffold, but also form significant portions of the tunnel wall (see Figure 19). The largest protein contributor to the surface of the tunnel is L22 whose long α-hairpin loop lies between RNA segments of domains I through IV and is approximately parallel with the axis of the tunnel. Unlike the other tunnel proteins, protein L39 does not have a globular domain at the surface of the particle and is almost entirely buried in domains I and III underneath protein L23. Interestingly, the nucleotides of 23S rRNA that form the tunnel wall are predominantly from loops in the 23S rRNA secondary structure (see Figure 19). As it progresses through the tunnel from the active site, a nascent polypeptide first encounters domain V followed 20 Å further along by domains II and IV and proteins L4 and L22. The last half of the tunnel is formed by domains I and III and the protein L39e.

The narrowest part of the tunnel is formed by proteins L22 and L4 which approach the tunnel from opposite sides forming what appears to be a gated opening (see Figure 19C). The function of this constriction, if any, is not obvious. It might be the place where the nature of the nascent chain is sensed and the information transmitted to the surface of the particle, perhaps through L22 or L4. The α-hairpin of L22 at the site of this orifice and the 23S rRNA interacting with it are highly conserved; its globular portion is located adjacent to the tunnel exit on the surface that must face the translocon during protein secretion (see Figure 19).

The "non-stick" character of the tunnel wall must reflect a lack of structural and polarity complementarity to any protein sequence or conformation that it encounters. The tunnel surface is largely hydrophilic and includes exposed hydrogen bonding groups from bases, backbone phosphates and polar protein side-chains (see Figure 19). While there are many hydrophobic groups (sugars, bases, protein side-chains) facing the tunnel as well, there are no patches of hydrophobic surface large enough to form a significant binding site for hydrophobic sequences in the nascent polypeptide. As the tunnel is some 20 Å in diameter and filled with water and the newly synthesized polypeptide is presumably freely mobile, the binding of a peptide to the tunnel wall would result in a large loss of entropy that would have to be compensated for by a large complementary interaction surface that is larger than 700 Å (Chothia *et al.* (1975) *Nature* 256: 705-708). Similarly, while Arg and Lys side-chains from a nascent peptide may indeed interact with the phosphates exposed in the tunnel, the degree of structural complementarity and the net binding energy obtained after displacing bound counterions must be too small to

overcome the large unfavorable entropy of immobilization that would result from peptide binding. Thus, although the ribosome tunnel is made primarily of RNA, the nature of its surface is reminiscent of the interior surface of the chaperonin, GroEL (Xu *et al.* (1998) *J. Struct. Biol.* 124:129-141) in its non-binding conformation. Only in the conformation that exposes a large hydrophobic surface does GroEL bind a denatured protein.

There are six proteins (L19, L22, L23, L24, L29 and L31e) located at the exit from the tunnel, facing the translocon onto which the ribosome docks during protein secretion. There is evidence that the ribosome binds the translocon even after extensive digestion of its protein by protease implying that interaction between the translocon and the ribosome is mediated by RNA. The proximity of these proteins to the translocon, however, leads us to wonder what role, if any, they might play in the protein secretion process. Recent data from the Dobberstein laboratory shows that the N-terminal domain of SRP54, the G-protein from the signal recognition particle involved in signal peptide binding, can be crosslinked to ribosomal proteins L23 and L29. These two proteins are adjacent to each other and at the tunnel exit (see Figure 19).

7. Evolution.

In vitro evolution of RNA oligonucleotides has produced small RNA molecules that can bind molecules like the Yarus inhibitor effectively or catalyze the peptidyl transfer reaction (Zhaug *et al.* (1998) *Chem. Biol.* 5: 539-553; Welch *et al.* (1997) *supra*). The sequence and secondary structure of one of these selected RNAs is reminiscent of the peptidyl transferase loop in domain V of 23S rRNA (Zhaug *et al.* (1998) *supra*). The most striking similarity is a five nucleotide sequence that is identical to a sequence in domain V that includes the catalytic A2486, G2482 and the buried phosphate of A2485. Remarkably, all of the groups involved in the proposed charge relay system for activating A2486 in the ribosome, are present in the *in vitro* selected ribozyme. Thus, though the surrounding structural context is likely to be different, it seems plausible that this artificially evolved ribozyme uses the same mechanisms as the ribosome for shifting the pKa of an adenine and likewise uses it as a base for peptide synthesis. A second RNA (Welch *et al.* (1997) *supra*) contains a 12 nucleotide loop that includes a 9-base sequence identical to that found in the same region of the peptidyl transferase loop.

The striking similarities between the sequences containing the key catalytic elements found in the peptidyl-transferase active site of the ribosome and sequences of *in vitro* selected

RNAs having related activities make it clear that the appearance of a small RNA domain capable of catalyzing peptidyl transferase was a plausible first step in the evolution of protein synthesis on the ribosome. The first peptides synthesized by this primordial peptide synthesizing enzyme might have been random polymers or copolymers, and it may have functioned with substrates as simple as an aminoacylated CCA. Basic peptides of the types observed to form the non-globular extensions that co-fold with the 23S rRNA might have been among the first peptides synthesized that were functionally useful. Such peptides may have enhanced the stability of the protoribosome and other early ribozymes as the more sophisticated peptides of the present day ribosome appear to do.

C. Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution, Complete Refinement

The three-dimensional structure of the large ribosomal subunit from *Haloarcula marismortui* has now been completely refined at 2.4Å resolution. The model includes 2876 RNA nucleotides, 3701 amino acids from 28 ribosomal proteins, 117 magnesium ions, 88 monovalent cations, and 7898 water molecules. Many of its proteins consist of a surface-exposed globular domain and one or more basic, non-globular extensions that are buried in rRNA. Half of them include motifs common in non-ribosomal proteins including, for example: RRM domains, SH3-like barrels and zinc fingers. Proteins that have significant sequence and structural similarity, such as L15 and L18e, make essentially identical interactions with rRNA.

More particularly, the *H. marismortui* 50S subunit has been completely rebuilt and refined by successive rounds of gradient energy minimization and B-factor refinement using CNS (Brünger *et al.* (1998) *supra*). Ribosomal proteins and rRNA were completely rebuilt using the software program "O" (Jones, T.A. *et al.* (1991) *Acta Crystallogr.* A46: 110-119) with $2F_o$-Fc electron density maps prior to the modeling of solvent and metal ions. Modeling errors in the proteins were identified using PROCHECK (Laskowski *et al.*(1993) *J. Appl. Cryst.* 26: 283-291) and by inspection of F_o-Fc maps. Difference maps also aided in the identification of errors in the rRNA, most often associated with sugar puckers. In the process, some adjustments were made in amino acid conformations, sequence register and in sequences themselves. Sequence changes made were largely limited to L10e, L15e, and L37Ae, the only three proteins from the *H. marismortui* 50S that have not been sequenced directly. In addition, fifty-one amino acids were

added to the model described in section IIA with forty-four of these coming from L10 at the base of the L7/L12 stalk and L39e which lines a portion of the wall of the polypeptide exit tunnel. Fewer adjustments were made to the rRNA structure. Forty-nine new nucleotides were modeled and refined, mainly in helices 43 and 44 in domain II of 23S rRNA. In addition, the sugar pucker or conformation about the glycosidic bond was adjusted for some nucleotides. The refinement process was monitored by the quality of electron density maps calculated using phases derived from the model as well as R / R_{free} values. The completely refined model now includes 2876 RNA nucleotides, 3701 amino acids, 210 metal ions, and 7898 water molecules. The model refines to an R/R_{free} of 18.9% / 22.3% and has excellent geometry (Table 4).

Solvent modeling began with the generation of a list of possible magnesium ions obtained by an automatic peak selection using CNS. Peaks greater than 3.5σ in F_o-Fc maps positioned within the magnesium inner-sphere coordination distance of 1.9 - 2.1 Å from N or O atoms were selected. The resulting list was manually inspected and only peaks that displayed clear octahedral coordination geometry were selected as magnesium.

Monovalent cations were identified on the basis of isomorphous differences between rubidium-soaked and native crystals of the *H. marismortui* 50S subunit. Since native crystals were stabilized in the presence of 1.6M NaCl, these sites were initially modeled and refined as Na^{+1}. Refinement of these sites as K^{+1} almost always resulted in unusually high temperature factors, with two exceptions where we have modeled K^{+1}. Most of the monovalent sites in the 50S subunit appear to occupied by Na^{+1} in our crystals, however, these sites are likely to be occupied by K^{+1} *in vivo*.

Waters were selected as peaks greater than 3.5σ in F_o-Fc electron density maps and between 2.5 and 3.3 Å of O or N atoms. Individual B-factor values were used to assess the assignment of water molecules. A number of waters refined to B-factors significantly lower than surrounding RNA and protein atoms. In many cases these peaks were found to be metal or chloride ions. A small number of low B-factor water molecules were retained in the final model because they could not be unambiguously assigned as other species. As a result of adding metal ions and water molecules the final model now contains 98, 547 non-hydrogen atoms. The refinement and model statistics for the large ribosomal subunit are summarized in Table 4.

Table 4
Refinement and Model Statistics for the *H. marismortui* 50S Subunit

Space Group	$C222_1$
a = 211.66Å, b = 299.67Å, c = 573.77Å	
Total non-hydrogen atoms	98,542
RNA atoms	61,617
Protein atoms	28,800
Water molecules	7,893
Magnesium ions	117
Potassium ions	2
Sodium ions	86
Chloride ions	22
Cadmium ions	5
Refinement Statistics:	
Resolution Range	15.0 - 2.4 Å
Number of reflections used in refinement	623,525
Number of reflections for cross-validation	6,187
$R_{working}$	18.9%
R_{free}	22.3%
σ_a coordinate error (cross-validated)	0.35Å (0.43Å)
luzzati coordinate error (cross-validated)	0.29Å (0.35Å)
Deviations from ideality:	
r.m.s.d. bond lengths	0.0052Å
r.m.s.d. bond angles	1.13°
r.m.s.d. dihedrals	15.7°
r.m.s.d. impropers	2.12°
Protein Statistics from Ramachandran Plot:	
Residues in most favored regions	2704 (86.6%)
Residues in additional allowed regions	379 (12.1%)
Residues in generously allowed regions	27 (0.9%)
Residues in disallowed regions	13 (0.4%)
Average B-factor Statistics ($Å^2$):	
All atoms (high / low)	44.3 (10.1 / 133.7)
rRNA	41.2 (11.78 / 125.0)
proteins	49.7 (13.9 / 92.5)
waters	41.89 (9.58 / 115.4)

Refinement has also permitted additional modeling of L10, L39e, and the L11 binding site in 23S rRNA. Furthermore, it has been discovered that certain motifs, for example, RRM topologies, SH3-like barrels and zinc fingers are common in the 50S proteins and each recognizes rRNA in many different ways. Proteins that have significant three-dimensional homology, however, such as L15 and L18e as well as L18 and S11, make essentially identical interactions with rRNA. Additional structural homologies between 50S proteins and non-ribosomal proteins also are apparent. The solvent exposed surfaces of these globular protein domains are rich in aspartate and glutamate residues, while irregular protein extensions penetrate

the RNA core of the ribosome. These extensions are often highly conserved, and their abundance of arginine, lysine, and glycine residues is important for their function. Collectively, the results show evolutionary connections between many ribosomal proteins and illustrate that protein-RNA interactions in the ribosome, although largely idiosyncratic, share some common principles.

D. Antibiotic Binding Sites

In addition to the foregoing structural studies, the structure of the large ribosomal subunit of *H. marismortui* has been determined complexed with each of seven different antibiotics. More specifically, crystals of the *H. marismortui* large ribosomal subunit have been soaked with one of the following antibiotics: anisomycin, blasticidin, carbomycin, tylosin, sparsomycin, virginiamycin or spiramycin. The structure of the large ribosome subunit complexed with each antibiotic was then resolved based on X-ray diffraction data generated for each crystal.

Briefly, a small amount of a concentrated antibiotic solution was added to a large subunit crystal suspended in stabilization solution and incubated for several hours. Following freezing and the other procedures normally used to prepare such crystals for experimental use, X-ray diffraction data were collected from the antibiotic containing crystals. Because the crystals were isomorphous to those from which the structure described above was derived, the phases obtained for native crystal were combined with the diffraction intensities obtained from the antibiotic-soaked crystal to obtain a structure for the latter. The position of the antibiotic in the crystal to which was bound is revealed most clearly in difference electron density maps, which are electron density maps computed using the phases just referred to and amplitudes obtained by subtracting the amplitudes of crystals that contain no antibiotic from the (suitably scaled) amplitudes of those that contain antibiotic. By using the foregoing methods, it was possible to determine the atomic co-ordinates that show the spatial relationship between particular antibiotics and their binding sites within the large ribosomal subunit. It is contemplated that similar methods can be used to resolve the structure of other antibiotics complexed to the large ribosomal subunit.

The atomic co-ordinates of the large ribosomal subunit complexed with anisomycin are listed in a table on compact disk Disk No. 3 of 3 under the file name anisomycin.pdb. In addition, Figure 20 shows the spatial relationship between the antibiotic anisomycin and the large ribosomal subunit.

The atomic co-ordinates of the large ribosomal subunit complexed with blasticidin are listed in a table on compact Disk No. 3 of 3 under the file name blasticidin.pdb. Figure 21 shows the spatial relationship between the antibiotic blasticidin and the large ribosomal subunit. For orientation, Figure 21 also includes a substrate for the P-site.

The atomic co-ordinates of the large ribosomal subunit complexed with carbomycin are listed in a table on compact disk Disk No. 3 of 3 under the file name carbomycin.pdb. Figure 22 shows the spatial relationship between the antibiotic carbomycin and the large ribosomal subunit. Figure 22 also shows a portion of the polypeptide exit tunnel.

The atomic co-ordinates of the large ribosomal subunit complexed with tylosin are listed in a table on compact disk Disk No. 3 of 3 under the file name tylosin.pdb. Figure 22 shows the spatial relationship between the antibiotic tylosin and the large ribosomal subunit. Figure 22 also shows a portion of the polypeptide exit tunnel.

The atomic co-ordinates of the large ribosomal subunit complexed with sparsomycin are listed in a table on compact disk Disk No. 3 of 3 under the file name sparsomycin.pdb. Figure 23 shows the spatial relationship between the antibiotic sparsomycin and the large ribosomal subunit. For orientation, Figure 23 also shows a substrate for the P-site.

The atomic co-ordinates of the large ribosomal subunit complexed with virginiamycin are listed in a table on compact disk Disk No. 3 of 3 under the file name virginiamycin.pdb. Figure 24 shows the spatial relationship between the antibiotics virginiamycin as well carbomycin, and the large ribosomal subunit.

The atomic co-ordinates of the large ribosomal subunit complexed with spiramycin are listed in a table on compact disk Disk No. 3 of 3 under the file name spiramycin.pdb.

Figure 25 shows the spatial orientations of several antibiotics, namely, blasticidin, anisomycin, virginiamycin and carbomycin, as they bind to their respective antibiotic binding sites within the large ribosomal subunit. For purposes of orienting the reader, the positions of the P-site, A-site and the polypeptide exit tunnel are shown in Figure 25. As is apparent, these antibiotics bind to or contact specific locations within the large ribosomal subunit to disrupt protein biosynthesis. For example, it appears that blasticidin binds the large ribosomal subunit in the vicinity of the P-site; anisomycin and virginiamycin bind the large ribosomal subunit in the

vicinity of the A-site; and carbomycin (a macrolide) binds the large ribosomal subunit in the vicinity of the polypeptide exit tunnel adjacent the peptidyl transferase site.

From Figure 25, it is apparent that the skilled artisan may identify certain portions of each antibiotic that contact regions in the large ribosomal subunit. By knowing their spatial relationship with respect one another, the skilled artisan may generate a hybrid antibiotic molecule comprising a portion of a first template antibiotic and a portion of a second, different template antibiotic. The two portions may be linked by a chemical linker so as to maintain the spatial orientation of one portion with respect to the other portion. As a result, the hybrid antibiotic may simultaneously bind each of the regions of the ribosomal subunit typically bound by each template antibiotic. The design and testing of such molecules is discussed in more detail below.

E. Experimental Techniques Which Exploit X-Ray Diffraction Data

Based on the X-ray diffraction pattern obtained from the assemblage of the molecules or atoms in a crystalline solid, the electron density of that solid may be reconstructed using tools well known to those skilled in the art of crystallography and X-ray diffraction techniques. Additional phase information extracted either from the diffraction data and available in the published literature and/or from supplementing experiments may then used to complete the reconstruction.

For basic concepts and procedures of collecting, analyzing, and utilizing X-ray diffraction data for the construction of electron densities see, for example, Campbell *et al.* (1984) Biological Spectroscopy, The Benjamin/Cummings Publishing Co., Inc., (Menlo Park, CA); Cantor *et al.* (1980) Biophysical Chemistry, Part II: Techniques for the study of biological structure and function, W.H. Freeman and Co., San Francisco, CA; A.T. Brünger (1993) X-PLOR Version 3.1: A system for X-ray crystallography and NMR, Yale Univ. Pr., (New Haven, CT); M.M. Woolfson (1997) An Introduction to X-ray Crystallography, Cambridge Univ. Pr., (Cambridge, UK); J. Drenth (1999) Principles of Protein X-ray Crystallography (Springer Advanced Texts in Chemistry), Springer Verlag; Berlin; Tsirelson *et al.* (1996) Electron Density and Bonding in Crystals: Principles, Theory and X-ray Diffraction Experiments in Solid State Physics and Chemistry, Inst. of Physics Pub.; U.S. Patent No. 5,942,428; U.S. Patent No.

6,037,117; U.S. Patent No. 5,200,910 and U.S. Patent No. 5,365,456 ("Method for Modeling the Electron Density of a Crystal").

A molecular model may then be progressively built using the experimental electron density information and further refined against the X-ray diffraction data resulting in an accurate molecular structure of the solid.

F. Structural Determination Of Other Large Ribosomal Subunits

It is understood that the skilled artisan, when provided with the atomic co-ordinates of a first macromolecule may use this information to quickly and easily determine the three-dimensional structure of a different but structurally related macromolecule. For example, the atomic co-ordinates defining the large ribosomal subunit from *H. marismortui* can be used to determine the structure of the large ribosomal subunit from other species either as an isolated subunit, in complex with the small subunit, or either of these complexed with functionally important ligands, for example: aminoacyl tRNA; various protein synthesis factors, such as elongation factor G, elongation factor Tu, termination factor or recycling factor, in both their GTP and GDP conformational states; and protein synthesis inhibitors, for example, antibiotics. In addition, the *H. marismortui* subunit co-ordinates can also be used to solve the structures of ribosomal complexes with components of the protein secretion machinery, for example, the signal recognition particle, and the translocon.

If the crystal being examined contains a macromolecule of unknown structure and no additional information is available, additional experiments sometimes may be required to determine the relevant phases of the macromolecule. These studies can often be time consuming and uncertain of success (Blundell *et al.* (1976) *supra*). However, when additional information, for example, structural and/or crystallographic information, is available for molecules related in some way to the macromolecule of interest then the process of resolving the structure of the molecule of interest is a much less challenging and time-consuming task.

Accordingly, the skilled artisan may use information gleaned from the prior resolved structure to develop a three-dimensional model of a new molecule of interest. Furthermore, the skilled artisan may use a variety of approaches to elucidate the three-dimensional structure of the new molecule. The approaches may depend on whether crystals of the molecule of interest are

available and/or whether the molecule of interest has a homologue whose structure has already been determined.

In one approach, if the molecule of interest forms crystals that are isomorphous, *i.e.*, that have the same unit cell dimensions and space group as a related molecule whose structure has been determined, then the phases and/or co-ordinates for the related molecule can be combined directly with newly observed amplitudes to obtain electron density maps and, consequently, atomic co-ordinates of the molecule of interest. The resulting maps and/or atomic co-ordinates may then be refined using standard refinement techniques known in the art. In another approach, if the molecule of interest is related to another molecule of known three-dimensional structure, but crystallizes in a different unit cell with different symmetry, the skilled artisan may use a technique known as molecular replacement to obtain useful phases from the co-ordinates of the molecule whose structure is known (Blundell *et al.* (1976) *supra*). These phases can then be used to generate an electron density map and/or atomic co-ordinates for the molecule of interest. In another approach, if no crystals are available for the molecule of interest but it is homologous to another molecule whose three-dimensional structure is known, the skilled artisan may use a process known as homology modeling to produce a three-dimensional model of the molecule of interest. It is contemplated that other approaches may be useful in deriving a three-dimensional model of a molecule of interest. Accordingly, information concerning the crystals and/or atomic co-ordinates of one molecule can greatly facilitate the determination of the structures of related molecules.

The method of molecular replacement, developed initially by Rossmann and Blow in the 1960s, is now used routinely to establish the crystal structures of macromolecules of unknown structures using the structure of a homologous molecule, or one in a different state of ligation (M.G. Rossmann, ed. "The Molecular Replacement Methods," *Int. Sci. Rev. J.* No. 13, Gordon & Breach, New York, NY (1972); Eaton Lattman, "Use of Rotation and Translation Functions," *H.W. Wyckoff, C.H.W. Hist.* (S.N. Timasheff, ed.) Methods in Enzymology, 115: 55-77 (1985)). For an example of the application of molecular replacement, see, for example, Rice, P.A. & Steitz, T.A. (1994) *EMBO J.* 13: 1514-24.

In molecular replacement, the three-dimensional structure of the known molecule is positioned within the unit cell of the new crystal by finding the orientation and position that

provides the best agreement between observed diffraction amplitudes and those calculated from the co-ordinates of the positioned subunit. From this modeling, approximate phases for the unknown crystal can be derived. In order to position a known structure in the unit cells of an unknown, but related structure, three rotation angles and three translations relative to the unit cell origin have to be determined. The rotation search is carried out by looking for agreement between the Patterson function of the search and target structures as a function of their relative orientation (the rotation function). X-PLOR (Brünger *et al.* (1987) *Science* 235:458-460; CNS (Crystallography & NMR System, Brünger *et al.*, (1998) *Acta Cryst. Sect. D* 54: 905-921), and AMORE: an Automatic Package for Molecular Replacement (Navaza, J. (1994) *Acta Cryst. Sect. A*, 50: 157-163) are computer programs that can execute rotation and translation function searches. Once the orientation of a test molecule is known, the position of the molecule must be found using a translational search. Once the known structure has been positioned in the unit cell of the unknown molecules, phases for the observed diffraction data can be calculated from the atomic co-ordinates of the structurally related atoms of the known molecules. By using the calculated phases and X-ray diffraction data for the unknown molecule, the skilled artisan can generate an electron density map and/or atomic co-ordinates of the molecule of interest.

By way of example, it is contemplated that a three-dimensional model of a ribosomal subunit other than that derived from *H. marismortui* can be generated via molecular replacement. In this method, the *H. marismortui* subunit structures are positioned within the unit cell of the new crystal by finding the orientation and position that provides the best agreement between observed diffraction amplitudes and those calculated from the co-ordinates of the positioned subunit. A starting electron density map calculated using $2F_{hkl}$(observed) - F_{hkl}(calculated), where F(observed) are the diffraction amplitudes that have been measured from crystals of the unknown structure, and F(calculated) are the diffraction amplitudes calculated from the positioned *H. marismortui* subunit structure. Refinement of the initial model can be done as is standard in the field of macromolecular crystallography.

The *H. marismortui* 50S structure can also be used to establish the structure of a 70S ribosome or 50S ribosome for which an election density map has been calculated, at a resolution that would otherwise be too low to be interpreted, while a 5 Å resolution map could not be interpreted in atomic terms *de novo*, a plausible atomic model can be constructed by refitting the

H. marismortui 50S structure to a lower resolution map (*e.g.*, 4.5 Å to 8 Å). This refitting can be combined with homology modeling to obtain a three-dimensional model of a ribosome or ribosomal subunit from a different species. It is contemplated that similar procedures may be used to determine the structure of the eukaryotic 60S subunit and/or a eukaryotic ribosome.

In general, the success of molecular replacement for solving structures depends on the fraction of the structures that are related and their degree of identity. For example, if about 50% or more of the structure shows an r.m.s. difference between corresponding atoms in the range of about 2 Å or less, the known structure can be successfully used to solve the unknown structure.

Homology modeling, also known as comparative modeling or knowledge-based modeling, can be used to generate a three-dimensional model for a molecule based on the known structure of homologues. In general, the procedure may comprise one or more of the following steps: aligning the amino acid or nucleic acid sequence of an unknown molecule against the amino acid or nucleic acid sequence of a molecule whose structure has previously been determined; identifying structurally conserved and structurally variable regions; generating atomic co-ordinates for core (structurally conserved) residues of the unknown structure from those of the known structure(s); generating conformations for the other (structurally variable) residues in the unknown structure; building side chain conformations; and refining and/or evaluating the unknown structure.

By way of example, since the nucleotide sequences of all known 50S subunit rRNAs can be aligned relative to each other and to *H. marismortui* 23S and 5S rRNAs, it is possible to construct models of the structures of other 50S ribosomal rRNAs, particularly in the regions of the tunnel and active sites, using the *H. marismortui* structure. Likewise, homologous proteins can also be modeled using similar methodologies. Methods useful for comparative RNA sequence analysis are known in the art and include visual methods and number pattern methods, as well as methods employing chi-square statistics, phylogenetic algorithms, or empirical algorithms. Descriptions of some of the foregoing methods are available, for example, at http://www.rna.icmb.utexas.edu/; Gutell (1996), "Comparative Sequence Analysis and the Structure of 16S and 23S rRNA," Ribosomal RNA. Structure, Evolution, Processing, and Function in Protein Biosynthesis, (Dahlberg A. and Zimmerman B., eds.) CRC Press. Boca Raton, pp. 111-128; Guttell *et al.* (1993) *Nucl. Acid Res.* 21: 3055 - 3074; Schnare *et al.* (1996) *J. Mol. Biol.* 256:

701-719. Particularly useful visual inspection methods include comparison of a particular position in a *H. marismortui* secondary structure diagram with the residues located at the analogous position on an *E. coli* secondary structure diagram. A software program that is particularly useful in homology modeling includes XALIGN (Wishart, D. *et al.*, (1994) *Cabios* 10: 687-88). See also, U.S. Patent No. 5,884,230.

To model the rRNA of a new species, bases of the *H marismortui* rRNA can be replaced, using a computer graphics program such as "O" (Jones *et al.*, (1991) *Acta Cryst. Sect. A,* 47: 110-119), by those of the homologous rRNA, where they differ. In many if not most cases the same orientation of the base will be appropriate. Insertions and deletions may be more difficult and speculative, but the rRNA forming the peptidyl transferase site and the portion of the tunnel closest to it is very highly conserved with essentially no insertions and deletions. Automated web-based homology modeling can be performed using, for example, the computer programs SWISS-MODEL available through Glaxo Wellcome Experimental Research in Geneva, Switzerland, and WHATIF available on EMBL servers.

For other descriptions of homology modeling, see, for example, Gutell R.R. (1996), *supra*; Gutell R.R., *et al.* (1993) *Nucleic Acids Res.* 21: 3055-3074; Schnare *et al.* (1996) *J. Mol. Biol.*, 256: 701-719; Blundell *et al.* (1987) *Nature* 326: 347-352; Fetrow and Bryant (1993) *Bio/Technology* 11:479-484; Greer (1991) *Methods in Enzymology* 202: 239-252; and Johnson *et al.* (1994) *Crit. Rev. Biochem. Mol. Biol.* 29:1-68. An example of homology modeling can be found, for example, in Szklarz G.D (1997) *Life Sci.* 61: 2507-2520.

As discussed earlier, the large ribosomal subunit from prokaryotes and eukaryotes are structurally conserved. The amino acid sequences of the large ribosomal subunit from prokaryotes and eukaryotes can be aligned due to the evolutionary conservation of amino acid residues that are important for three-dimensional structure, the nature and shape of the binding sites for substrates and the catalytic site. This similarity in amino acid sequence of the homologous large ribosomal subunit allows the construction of models, via homology modeling, for the molecules whose crystal structures have not been solved.

The new ribosome or large ribosomal subunit structures determined using the *H. marismortui* crystals and/or atomic co-ordinates can then be used for structure-based drug design using one or more of the approaches described hereinbelow. This information can then be used

to design molecules that selectively bind and disrupt protein synthesis in the ribosomes of the pathogens while leaving the ribosomes of a host relatively unaffected.

G. Rational Drug Design

1. Introduction

It is contemplated that the atomic co-ordinates defining a large ribosomal subunit of interest, whether derived from one or more of X-ray crystallography, molecular modeling, homology modeling or molecular replacement, may be used in rational drug design (RDD) to design a novel molecule of interest, for example, novel modulators (for example, inducers, mimetics or inhibitors) of ribosome function. Furthermore, it is contemplated that, by using the principles disclosed herein, the skilled artisan can design, make, test, refine and use novel protein synthesis inhibitors specifically engineered to reduce, disrupt, or otherwise or inhibit ribosomal function in an organism or species of interest. For example, by using the principles discussed herein, the skilled artisan can engineer new molecules that specifically target and inhibit ribosomal function in a pathogen, for example, a particular prokaryotic, organism, while preserving ribosomal function in a host, for example, a eukaryotic organism, specifically a mammal, and more specifically, a human. As a result, the atomic co-ordinates provided and discussed herein permit the skilled artisan to design new antibiotics that can kill certain pathogenic organisms while having little or no toxicity in the intended recipient, for example, a human.

It is contemplated that RDD using atomic co-ordinates of the large ribosomal subunit can be facilitated most readily via computer-assisted drug design (CADD) using conventional computer hardware and software known and used in the art. The candidate molecules may be designed *de novo* or may be designed as a modified version of an already existing molecule, for example, a pre-existing antibiotic, using conventional methodologies. Once designed, candidate molecules can be synthesized using standard methodologies known and used in the art. Following synthesis, the candidate molecules can be screened for bioactivity, for example, by their ability to reduce or inhibit ribosome function, their ability to interact with or bind a ribosome or a ribosomal subunit. Based in part upon these results, the candidate molecules may be refined iteratively using one or more of the foregoing steps to produce a more desirable molecule with a desired biological activity. The resulting molecules can be useful in treating,

inhibiting or preventing the biological activities of target organisms, thereby killing the organism or impeding its growth. Alternatively, the resulting molecules can be useful for treating, inhibiting or preventing microbial infections in any organism, particularly animals, more particularly humans.

In summary, the tools and methodologies provided by the present invention may be used to identify and/or design molecules which bind and/or interact in desirable ways with ribosomes and ribosomal subunits. Basically, the procedures utilize an iterative process whereby the molecules are synthesized, tested and characterized. New molecules can be designed based on the information gained in the testing and characterization of the initial molecules and then such newly identified molecules can themselves be tested and characterized. This series of processes may be repeated as many times as necessary to obtain molecules with desirable binding properties and/or biological activities. Methods for identifying candidate molecules are discussed in more detail below.

2. *Identification of Candidate Molecules*

It is contemplated that the design of candidate molecules of interest can be facilitated by conventional ball and stick-type modeling procedures. However, in view of the size and complexity of the large ribosomal subunit, it is contemplated that the ability to design candidate molecules may be enhanced significantly using computer-based modeling and design protocols.

a. **Molecular Modeling.**

It is contemplated that the design of candidate molecules, as discussed in detail hereinbelow, can be facilitated using conventional computers or workstations, available commercially from, for example, Silicon Graphics Inc. and Sun Microsystems, running, for example, UNIX based, Windows NT on IBM OS/2 operating systems, and capable of running conventional computer programs for molecular modeling and rational drug design.

It is understood that any computer system having the overall characteristics set forth in Figure 27 may be useful in the practice of the invention. More specifically, Figure 27, is a schematic representation of a typical computer work station having in electrical communication (100) with one another via, for example, an internal bus or external network, a central processing unit (101), a random access memory (RAM) (102), a read only memory (ROM) (103), a monitor

or terminal (104), and optimally an external storage device, for example, a diskette, CD ROM, or magnetic tape (105).

The computer-based systems of the invention preferably comprise a data storage means having stored therein a ribosome or ribosomal subunit or fragment sequence and/or atomic co-ordinate/X-ray diffraction data of the present invention and the necessary hardware means and software means for supporting and implementing an analysis means. As used herein, "a computer system" or "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the sequence, X-ray diffraction data, and/or atomic co-ordinates of the invention. As used herein, the term "data storage means" is understood to refer to any memory which can store sequence data, atomic co-ordinates, and/or X-ray diffraction data, or a memory access means which can access manufactures having recorded thereon the atomic co-ordinates of the present invention.

In one embodiment, a ribosome or ribosomal subunit, or at least a subdomain thereof, amino acid and nucleic acid sequence, X-ray diffraction data and/or atomic co-ordinates of the present invention are recorded on computer readable medium. As used herein, the term "computer readable medium" is understood to mean any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as optical discs or CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon an amino acid and/or nucleotide sequence, X-ray diffraction data, and/or atomic co-ordinates of the present invention.

As used herein, the term "recorded" is understood to mean any process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising an amino acid or nucleotide sequence, atomic co-ordinates and/or X-ray diffraction data of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon amino acid and/or nucleotide sequence,

atomic co-ordinates and/or X-ray diffraction data of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the sequence information, X-ray data and/or atomic co-ordinates of the present invention on computer readable medium. The foregoing information, data and co-ordinates can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MICROSOFT Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the information of the present invention.

By providing a computer readable medium having stored thereon a ribosome or ribosomal subunit sequence, and/or atomic co-ordinates, a skilled artisan can routinely access the sequence, and/or atomic co-ordinates to model a ribosome or ribosomal subunit, a subdomain thereof, mimetic, or a ligand thereof. Computer algorithms are publicly and commercially available which allow a skilled artisan to access this data provided in a computer readable medium and analyze it for molecular modeling and/or RDD. See, *e.g.*, Biotechnology Software Directory, MaryAnn Liebert Publ., New York, NY (1995).

Although computers are not required, molecular modeling can be most readily facilitated by using computers to build realistic models of a ribosome, ribosomal subunit, or a portion thereof. Molecular modeling also permits the modeling of new smaller molecules, for example ligands, agents and other molecules, that can bind to a ribosome, ribosomal subunit, or a portion therein. The methods utilized in molecular modeling range from molecular graphics (*i.e.*, three-dimensional representations) to computational chemistry (*i.e.*, calculations of the physical and chemical properties) to make predictions about the binding of the smaller molecules or their activities; to design new molecules; and to predict novel molecules, including ligands such as drugs, for chemical synthesis.

For basic information on molecular modeling, see, for example, M. Schlecht, Molecular Modeling on the PC (1998) John Wiley & Sons; Gans *et al.*, Fundamental Principals of Molecular Modeling (1996) Plenum Pub. Corp.; N.C. Cohen, ed., Guidebook on Molecular

Modeling in Drug Design (1996) Academic Press; and W.B. Smith, Introduction to Theoretical Organic Chemistry and Molecular Modeling (1996). U.S. Patents which provide detailed information on molecular modeling include, for example: U.S. Patent Nos. 6,093,573; 6,080,576; 6,075,014; 6,075,123; 6,071,700; 5,994,503; 5,884,230; 5,612,894; 5,583,973; 5,030,103; 4,906,122; and 4,812,12.

Three-dimensional modeling can include, but is not limited to, making three-dimensional representations of structures, drawing pictures of structures, building physical models of structures, and determining the structures of related ribosomes, ribosomal subunits and ribosome/ligand and ribosomal subunit/ligand complexes using the known co-ordinates. The appropriate co-ordinates are entered into one or more computer programs for molecular modeling, as known in the art. By way of illustration, a list of computer programs useful for viewing or manipulating three-dimensional structures include: Midas (University of California, San Francisco); MidasPlus (University of California, San Francisco); MOIL (University of Illinois); Yummie (Yale University); Sybyl (Tripos, Inc.); Insight/Discover (Biosym Technologies); MacroModel (Columbia University); Quanta (Molecular Simulations, Inc.); Cerius (Molecular Simulations, Inc.); Alchemy (Tripos, Inc.); LabVision (Tripos, Inc.); Rasmol (Glaxo Research and Development); Ribbon (University of Alabama); NAOMI (Oxford University); Explorer Eyechem (Silicon Graphics, Inc.); Univision (Cray Research); Molscript (Uppsala University); Chem-3D (Cambridge Scientific); Chain (Baylor College of Medicine); O (Uppsala University); GRASP (Columbia University); X-Plor (Molecular Simulations, Inc.; Yale University); Spartan (Wavefunction, Inc.); Catalyst (Molecular Simulations, Inc.); Molcadd (Tripos, Inc.); VMD (University of Illinois/Beckman Institute); Sculpt (Interactive Simulations, Inc.); Procheck (Brookhaven National Library); DGEOM (QCPE); RE_VIEW (Brunell University); Modeller (Birbeck College, University of London); Xmol (Minnesota Supercomputing Center); Protein Expert (Cambridge Scientific); HyperChem (Hypercube); MD Display (University of Washington); PKB (National Center for Biotechnology Information, NIH); ChemX (Chemical Design, Ltd.); Cameleon (Oxford Molecular, Inc.); and Iditis (Oxford Molecular, Inc.).

One approach to RDD is to search for known molecular structures that might bind to a site of interest. Using molecular modeling, RDD programs can look at a range of different molecular structures of molecules that may fit into a site of interest, and by moving them on the

computer screen or via computation it can be decided which structures actually fit the site well (William Bains (1998) Biotechnology from A to Z, second edition, Oxford University Press, p. 259).

An alternative but related approach starts with the known structure of a complex with a small molecule ligand and models modifications of that small molecule in an effort to make additional favorable interactions with a ribosome or ribosomal subunit.

The present invention permits the use of molecular and computer modeling techniques to design and select novel molecules, such as antibiotics or other therapeutic agents, that interact with ribosomes and ribosomal subunits. Such antibiotics and other types of therapeutic agents include, but are not limited to, antifungals, antivirals, antibacterials, insecticides, herbicides, miticides, rodentcides, etc.

In order to facilitate molecular modeling and/or RDD the skilled artisan may use some or all of the atomic co-ordinates deposited at the RCSB Protein Data Bank with the accession numbers PDB ID: 1FFK, 1JJ2, 1FFZ, or IFG0, and/or those atomic co-ordinates contained on Disk No. 1, 2 or 3 of 3. Furthermore, the skilled artisan, using the foregoing atomic co-ordinates, the skilled artisan can generate additional atomic co-ordinates via, for example, molecular modeling using, for example, homology modeling and/or molecular replacement techniques, that together define at least a portion of a model of a ribosome from another species of interest. By using the foregoing atomic co-ordinates, the skilled artisan can design inhibitors of protein synthesis that may be tailored to be effective against ribosomes from one or more species but which have little or no effect on ribosomes of other species. Such inhibitors may be competitive inhibitors. As used herein, the term "competitive inhibitor" refers to an inhibitor that binds to the active form of a ribosome or ribosomal subunit at the same sites as its substrate(s) or tRNA(s), thus directly competing with them. The term "active form" of a ribosome or ribosomal subunit refers to a ribosome or ribosomal subunit in a state that renders it capable of protein synthesis. Competitive inhibition can be reversed completely by increasing the substrate or tRNA concentration.

This invention also permits the design of molecules that act as uncompetitive inhibitors of protein synthesis. As used herein, the term "uncompetitive inhibitor" refers to a molecule that inhibits the functional activity of a ribosome or ribosomal subunit by binding to a different site

on the ribosome or ribosomal subunit than does its substrates, or tRNA. Such inhibitors can often bind to the ribosome or ribosomal subunit with the substrate or tRNA and not to the ribosome or ribosomal subunit by itself. Uncompetitive inhibition cannot be reversed completely by increasing the substrate concentration. These inhibitors may bind to, all or a portion of, the active sites or other regions of the large ribosomal subunit already bound to its substrate and may be more potent and less non-specific than known competitive inhibitors that compete for large ribosomal subunit active sites or for binding to large ribosomal subunit.

Similarly, non-competitive inhibitors that bind to and inhibit protein synthesis whether or not it is bound to another chemical entity may be designed using the atomic co-ordinates of the large ribosomal subunits or complexes comprising large ribosomal subunit of this invention. As used herein, the term "non-competitive inhibitor" refers to an inhibitor that can bind to either the free or substrate or tRNA bound form of the ribosome or ribosomal subunit.

Those of skill in the art may identify inhibitors as competitive, uncompetitive, or non-competitive by computer fitting enzyme kinetic data using standard equation according to Segel, I.H., (1975) Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems, (Wiley Classics Library). It should also be understood that uncompetitive or non-competitive inhibitors according to the present invention may bind the same or different binding sites.

Alternatively, the atomic co-ordinates provided by the present invention are useful in designing improved analogues of known protein synthesis inhibitors or to design novel classes of inhibitors based on the atomic structures and co-ordinates of the crystals of the 50S ribosomal subunit/CCdA-p-Puro complex and the 50S ribosomal subunit/aa-tRNA analogue complex. This provides a novel route for designing inhibitors of protein synthesis with both high specificity, stability and other drug-like qualities (Lipinski *et al.* (1997) *Adv. Drug Deliv. Rev.* 23:3).

The atomic co-ordinates of the present invention also permit probing the three-dimensional structure of a ribosome or ribosome subunit or a portion thereof with molecules composed of a variety of different chemical features to determine optimal sites for interaction between candidate inhibitors and/or activators and the ribosome or ribosomal subunit. For example, high resolution atomic co-ordinates based on X-ray diffraction data collected from crystals saturated with solvent allows the determination of where each type of solvent molecule

sticks. Small molecules that bind to those sites can then be designed and synthesized and tested for their inhibitory activity (Travis, J. (1993) *Science* 262: 1374). Further, any known antibiotic, inhibitor or other small molecule that binds to the *H. marismortui* large subunit can be soaked into *H. marismortui* large subunit crystals and their exact mode of binding determined from difference electron density maps. These molecules may represent lead compounds from which better drug-like compounds can be synthesized.

b. Identification of Target Sites.

The atomic co-ordinates of the invention permit the skilled artisan to identify target locations in a ribosome or large ribosomal subunit that can serve as a starting point in rational drug design. As a threshold matter, the atomic co-ordinates of the invention permit the skilled artisan to identify specific regions within a ribosome or ribosomal subunit that are involved with protein synthesis and/or protein secretion out of the ribosome. Furthermore, the atomic co-ordinates of the invention permit a skilled artisan to further identify portions of these regions that are conserved or are not conserved between different organisms. For example, by identifying portions of these regions that are conserved among certain pathogens, for example, certain prokaryotes, but are not conserved in a host organism, for example, a eukaryote, more preferably a mammal, the skilled artisan can design molecules that selectively inhibit or disrupt protein synthesis activity of the pathogen's but not the host's ribosomes. Furthermore, by analyzing regions that are either conserved or non-conserved between certain pathogens, it may be possible to design broad or narrow spectrum protein synthesis inhibitors, *e.g.*, antibiotics, as a particular necessity arises.

Figure 28, is a schematic representation of a large ribosomal subunit that identifies a variety of exemplary target sites that appear to participate in protein synthesis within the ribosome and/or the export or translocation of the newly synthesized protein out of the ribosome. The target sites include, for example, the P-site (200), the A-site (201), the peptidyl transferase center (202), the peptidyl transferase site (203) which includes at least a portion of the P-site and the A-site, a factor binding domain (204) including, for example, the EF-Tu binding domain and the EF-G binding domain, the polypeptide exit tunnel (205) including cavities defined by the wall of the exit tunnel, and the signal recognition particle binding domain (206).

By way of example, inspection of the atomic co-ordinates of the *H. marismortui* 50S ribosomal subunit has identified a variety of target regions that may serve as a basis for the rational drug design of new or modified protein synthesis inhibitors. The target regions include the peptidyl transferase site, A-site, the P-site, the polypeptide exit tunnel, certain cavities disposed in the wall of the polypeptide exit tunnel (for example, cavity 1 and cavity 2), and certain antibiotic binding pockets. The residues that together define at least a portion of each of the foregoing regions are identified in the following tables. However, it is contemplated that the same or similar target sites can be identified in a ribosome or a ribosomal unit of interest using the principles described herein. Furthermore, these principles can be employed using any of the primary sets of atomic co-ordinates provided herein or any additional atomic co-ordinate sets, for example, secondary atomic co-ordinate sets that may be generated by molecular modeling of any ribosome or ribosomal subunit of interest.

Table 5 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of the ribosomal peptidyl transferase site. In addition, Table 5 identifies which of those residues that define at least a portion of the peptidyl transferase site are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non-conserved residues were identified by comparison of sequences of *H. marismortui* 23S rRNA that form the above-mentioned sites with the corresponding sequences of aligned rRNA from the other organisms.

Table 5

Residues that Define the Ribosomal Peptidyl-Transferase Site

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
G2102				Yes
A2103				Yes
C2104				Yes
C2105				Yes
C2106				Yes
G2482				Yes
G2284				Yes
G2285				Yes
G2286				Yes
A2474				Yes
A2485				Yes
A2486				Yes
C2487				Yes
A2488		U	A/U	No
U2528				Yes
C2536				Yes
A2538				Yes
G2540				Yes
U2541				Yes
C2542				Yes
G2543				Yes
G2588				Yes
U2589				Yes
U2590				Yes
C2608				Yes
G2617				Yes
G2618				Yes
U2619				Yes
U2620				Yes
A2635				Yes
C2636				Yes
A2637				Yes
G2638				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of the CCdA-p-Puromycin A-site ligand and CCdA-Puromycin transition state inhibitor (PDB accession codes: 1fg0 and 1ffz, respectively) and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 6 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of the ribosomal A-site. In addition, Table 6 identifies which of those residues that define at least a portion the A-site are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non conserved residues were identified as described previously with respect to Table 5.

Table 6
Residues that Define the Ribosomal A-site

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
G2102				Yes
A2103				Yes
C2104				Yes
G2482				Yes
A2485				Yes
A2486				Yes
C2487				Yes
A2488		U	A/U	No
U2528				Yes
C2536				Yes
A2538				Yes
G2540				Yes
U2541				Yes
C2542				Yes
G2543				Yes
G2588				Yes
U2589				Yes
U2590				Yes
C2608				Yes
G2617				Yes
G2618				Yes
U2619				Yes
U2620				Yes
A2637				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of the CC-Puromycin A-site ligand and the 50S ribosome (PDB accession code 1fg0) using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 7 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of the ribosomal P-site. As demonstrated in Table 7, all of the residues in that portion of the ribosomal P-sites are conserved between *H. marismortui* and *E. coli*, between *H. marismortui* and rat, between *E. coli* and rat, and between eubacteria and eukaryota, as determined using the comparison method described previously with respect to Table 5.

Table 7
Residues that Define the Ribosomal P-site

H. marismortui P-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C2104				Yes
C2105				Yes
C2106				Yes
G2284				Yes
G2285				Yes
G2286				Yes
A2474				Yes
A2485				Yes
A2486				Yes
U2619				Yes
U2620				Yes
A2635				Yes
C2636				Yes
A2637				Yes
G2638				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of the CCA-PO_2 moiety of the CCdA-p-puromycin transition state inhibitor and the 50S ribosome (PDB accession code 1ffz) using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 8 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of the ribosomal polypeptide exit tunnel. In addition, Table 8 identifies which of those residues that define at least a portion of the polypeptide exit tunnel are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not

conserved between eubacteria and eukaryota. The non conserved residues were identified as described previously with respect to Table 5.

Table 8
Residues that Define the Ribosomal Peptide Exit Tunnel

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
23S rRNA				
G23				Yes
G24				Yes
A60				Yes
G88				Yes
G89				Yes
A90				Yes
U454				Yes
A462	C		C/A	No
A466				Yes
G467	C	A	C/A	No
U468	C	C		Yes
G469				Yes
A476				Yes
A477				Yes
C478	A	A		Yes
G487				Yes
U488	A	A		Yes
A489				Yes
C490				Yes
C491		U	C/U	No
C492		U	C/U	No
A497				Yes
A498		G	A/G	No
G499		A	G/A	No
G500				
G501		A	G/A	No
A513		G	A/G	No
G514	A		A/G	No
G636		A	G/A	No
A767				Yes
U768	A	A		Yes
U835	del		del/U	No
C839	U	G	U/G	No
U840				Yes
A841	G	G		Yes
A844				Yes
U845	A	A		Yes
C879				Yes
A882				Yes
U883				Yes
C884				Yes
G885	A	C	A/C	No

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
U888	C	C		Yes
C889				Yes
C890	G	A	G/A	No
U1359				Yes
C1360	G		G/C	No
C1361		U	C/U	No
U1362				Yes
G1363				Yes
G1364	A		A/G	No
C1365		U	C/U	No
C1366	A	G	A/G	No
A1367	U	G	U/G	No
A1424	U	G	U/G	No
G1425	C	U	C/U	No
C1426		G	C/G	No
A1427				Yes
C1428	A	A		Yes
U1429	C	C		Yes
G1430		A	G/A	No
C1439	G	A	G/A	No
U1440	G	C	G/C	No
G1441	C	A	C/A	No
A1442		U	A/U	No
A1689		C	A/C	No
C1690		A	C/A	No
A1691		G	A/G	No
A1836				Yes
G1837	U	A	U/A	No
U1838				Yes
A2054				Yes
A2055				Yes
C2056	A	A		Yes
U2057		C	U/C	No
G2058	U	C	U/C	No
C2098	G	A	G/A	No
G2099	A		A/G	No
A2100				Yes
A2101				Yes
G2102				Yes
A2103				Yes
C2104				Yes
C2105				Yes
A2474				Yes
C2475				Yes
C2476	U		U/C	No
C2477		A	C/A	No
U2478	C	C		Yes
G2482				Yes
A2485				Yes
A2486				Yes

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C2487				Yes
A2488		U	A/U	No
C2536				Yes
A2538				Yes
G2540				Yes
U2541				Yes
C2542				Yes
C2608				Yes
G2611				Yes
G2616				Yes
G2618				Yes
U2619				Yes
U2620				Yes
U2621				Yes
A2637				Yes
G2643				Yes
C2644	U	U		Yes
U2645	C		C/U	No
G2646	C	U	C/U	No
C2647		U	C/U	No
Protein L4				
E59				Yes
S60	V	S	V/S	No
F61	T	W	T/W	No
G62				Yes
S63		T	S/T	No
G64				Yes
R65	K		K/R	No
G66	DEL	A	DEL/A	No
Q67	DEL	V	DEL/V	No
A68	DEL		DEL/A	No
H69	K	R	K/R	No
V70	P	I	P/I	No
P71	W	P	W/P	No
K72	R	R		Yes
L73	K	R	K/R	No
D74	G	G		Yes
G75	T	G	T/G	No
R76	G	G		Yes
A77	R	T	R/T	No
Protein L22				
E20	H	N	H/N	No
E121	S	K	S/K	No
Q122	M	M		Yes
Q123	K	R	K/R	No
G124	R	R		Yes
R125	I	R	I/R	No

H. marismortui A-site Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
K126	M	T	M/T	No
P127		Y	P/Y	No
R128				Yes
A129				Yes
M130	K	H	K/H	No
G131				Yes
R132				Yes
A133		I	A/I	No
S134	D	N	D/N	No
A135	R	P	R/P	No
W136	I	Y	I/Y	No
N137	L	M	L/M	No
Q140	T	P	T/P	No
Protein L39E **				
N18				
S19				
R20				
V21				
P22				
A23				
Y24				
V25				
M26				
L27				
K28				
T29				
E31				
R35				
N36				
H37				
K38				
R39				
R40				
H41				
R44				
N45				

**Given that homologues for the H. marismortui ribosomal protein L39E are not currently known, sequence comparisons with *E. coli*, and *Rattus norvegicus* are not possible.

Residues were determined by a 10 angstrom distance measurement between the atoms of a model of a newly synthesized peptide positioned in the center of the exit tunnel and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/

c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Figure 26 shows a region of the large ribosomal subunit in which an antibiotic binds. Figure 26(A) shows an enlarged portion of the large ribosomal subunit with the antibiotic tylosin bound at the top of the polypeptide exit tunnel adjacent the peptidyl transferase site. Figures 26(B) and 26(C) are views showing each half of a large ribosomal subunit cut along the polypeptide exit tunnel and are provided to orient the reader to show the tylosin binding site relative to the large ribosomal unit as a whole. Figure 26(A) also shows two cavities defined by the wall of the polypeptide exit tunnel and are denoted as "cavity 1" and "cavity 2." In addition, Figure 26(A) also shows a disaccharide binding pocket. The direction in which the newly synthesized polypeptide chains exits the ribosome through the polypeptide exit tunnel is denoted by an arrow.

Table 9 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define a first cavity within the wall of polypeptide exit tunnel (cavity 1). In addition, Table 9 identifies which of those residues that define cavity 1 are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non-conserved residues were identified as described previously with respect to Table 5.

Table 9
Residues that Define Cavity 1 in the Ribosomal Peptide Exit Tunnel

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C474	G	G		Yes
A766				Yes
A767			A/C	Yes
U768	A	C		No
U883				Yes
C884			A/G	Yes
G885	A			No
A886				Yes
U888	C	C		Yes
C889			G/A	Yes
C890	G	A		No
U1359			U/A	Yes
G1837	U	A		No
A2100				Yes
A2101				Yes
G2102				Yes
A2103				Yes
C2475			U/C	No
C2476	U	A	C/A	No
C2477		C		Yes
U2478	C		G/A	No
A2479	G			Yes
A2538				
Protein L4			R/S	No
P57	R	S		Yes
A58				Yes
E59			V/S	No
S60	V		T/W	No
F61	T	W		Yes
G62			S/T	No
S63		T		Yes
G64			K/R	No
R65	K		DEL/V	No
Q67	DEL	V	P/I	No
V70	P	I	W/P	No
P71	W			Yes
K72	R	R	K/R	No
L73	K	R		Yes
D74	G	G	T/G	No
G75	T			Yes
R76	G	G		

Cavity residues were identified using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 10 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define a second cavity in the wall of polypeptide exit tunnel (cavity 2). In addition, Table 10 identifies which of those residues that define cavity 2 are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non conserved residues were identified as described previously with respect to Table 5.

Table 10
Residues that Define Cavity 2 in the Ribosomal Peptide Exit Tunnel

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
U831	C	G	C/G	No
U832	DEL		DEL/U	No
G833	U	C	U/C	No
G834	A	C	A/C	No
U835	DEL		DEL/U	No
G836	A		A/G	No
U837		A	U/A	No
C838	G		G/C	No
C839	U	G	U/G	No
U840				Yes
A841	G	G		Yes
A843				Yes
A844				Yes
U845	A	A		Yes
A846		U	A/U	No
C847	U		U/C	No
C848	U	G	U/G	No
C849	A	G	A/G	No
C1753				Yes
A1754				Yes
G1837	U	A	U/A	No
U1838				Yes
A1839				Yes
G2099	A		A/G	No

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
A2100				Yes
A2103				Yes
U2615				Yes
G2616				Yes
U2621				Yes
A2622				Yes
G2643				Yes
C2644	U	U		Yes
U2645	C		C/U	No
G2646	C	U	C/U	No
C2647		U	C/U	No

Cavity residues were determined using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Tables 9 and 10, however, define only two or many cavities disposed within the wall of the polypeptide exit tunnel. However, by using the atomic co-ordinates and molecular modeling methodologies described herein, the skilled artisan may identify the residues (contributed by amino acids, nucleotides or a combination of both) that together define other cavities within the wall of the polypeptide exit tunnel.

In addition, by using the atomic co-ordinates described herein, the skilled artisan can identify the antibiotic binding site of any antibiotic of interest. This information also provides contact sites between an antibiotic and the residues in a ribosome or ribosomal subunit, which can be used to advantage in the design of novel or modified protein synthesis inhibitors. The binding or contact sites for a variety of antibiotics are discussed in more detail below.

Table 11 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of an anisomycin binding pocket. In addition, Table 11 identifies which of those residues that define at least a portion of the anisomycin binding pocket are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not

conserved between eubacteria and eukaryota. The non-conserved residues were identified as described previously with respect to Table 5.

Table 11
Residues that Define the Anisomycin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
G2102				Yes
G2482				Yes
A2486				Yes
C2487				Yes
A2488		U	A/U	No
U2535				Yes
A2538				Yes
U2539				Yes
G2540				Yes
U2541				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of anisomysin and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 12 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a blasticidin binding pocket. As demonstrated in Table 12, all of the residues in that portion of the blasticidin binding pocket are conserved between *H. marismortui* and *E. coli*, between *H. marismortui* and rat, between *E. coli* and rat, and between eubacteria and eukaryota, as determined using the comparison method described previously with respect to Table 5.

Table 12
Residues that Define the Blasticidin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C2104				Yes
C2105				Yes
C2106				Yes
G2284				Yes
G2285				Yes
U2473				Yes
A2474				Yes
A2485				Yes
A2486				Yes
U2620				Yes
G2634				Yes
A2635				Yes
C2636				Yes
A2637				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of Blasticidin and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 13 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a carbomycin binding pocket. In addition, Table 13 identifies which of those residues that define at least a portion of the carbomycin binding pocket are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non-conserved residues were identified as described previously with respect to Table 5.

Table 13
Residues that Define the Carbomycin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C839	U	G	U/G	No
G2099	A		A/G	No
A2100				Yes
G2102				Yes
A2103				Yes
C2104				Yes
A2486				Yes
C2487				Yes
A2538				Yes
G2540				Yes
U2541				Yes
U2620				Yes
C2644	U	U		Yes
G2646	C	U	C/U	No

Residues were determined by a 5.8 angstrom distance measurement between the atoms of carbomycin and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 14 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a tylosin binding pocket. In addition, Table 14 identifies which of those residues that define at least a portion of the tylosin binding pocket are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat, and those that are not conserved between eubacteria and eukaryota. The non-conserved residues were identified as described previously with respect to Table 5.

Table 14
Residues that Define the Tylosin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C839	U	G	U/G	No
A841	G	G		Yes
A843				Yes
A844				Yes
U845	A	A		Yes
G1837	U	A	U/A	No
C2098	G	A	G/A	No
G2099	A		A/G	No
A2100				Yes
G2102				Yes
A2103				Yes
A2538				Yes
U2539				Yes
G2540				Yes
U2541				Yes
U2645	C		C/U	No
G2646	C	U	C/U	No

Residues were determined by a 5.8 angstrom distance measurement between the atoms of tylosin and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 15 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a sparsomycin binding pocket. As demonstrated in Table 15, all of the residues in that portion of the sparsomycin binding pocket are conserved between *H. marismortui* and *E. coli*, between *H. marismortui* and rat, between *E. coli* and rat, and between eubacteria and eukaryota, as determined using the comparison method described previously with respect to Table 5.

Table 15
Residues that Define the Sparsomycin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
A2486				Yes
C2487				Yes
U2541				Yes
C2608				Yes
U2619				Yes
U2620				Yes
C2636				Yes
A2637				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of sparsomycin and the 50S ribosome using
the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and Rattus *norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 16 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a virginiamycin binding pocket. As demonstrated in Table 16, all of the residues in that portion of the virginiamycin binding pocket are conserved between *H. marismortui* and *E. coli*, between *H. marismortui* and rat, between *E. coli* and rat, and between eubacteria and eukaryota, as determined using the comparison method described previously with respect to Table 5.

Table 16
Residues that Define the Virginiamycin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
A2100				Yes
G2102				Yes
A2103				Yes
C2104				Yes
C2105				Yes
G2482				Yes
A2486				Yes
C2487				Yes
U2535				Yes
C2536				Yes
A2538				Yes
U2539				Yes
G2540				Yes
U2541				Yes
U2620				Yes

Residues were determined by a 5.8 angstrom distance measurement between the atoms of virginiamycin and the 50S ribosome using the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

Table 17 identifies the residues in the *H. marismortui* 50S ribosomal subunit that together define at least a portion of a spiramycin binding pocket. In addition, Table 17 identifies which of those residues that define at least a portion of the spiramycin binding pocket are not conserved between *H. marismortui* and *E. coli*, those that are not conserved between *H. marismortui* and rat, those that are not conserved between *E. coli* and rat and those that are not conserved between eubacteria and eukaryota. The non-conserved residues were identified as described previously with respect to Table 5.

Table 17
Residues that Define the Spiramycin Binding Pocket

H. marismortui Residues	Residues Not Conserved between H. marismortui and E. coli	Residues Not Conserved Between H. marismortui and Rattus	Residues Not Conserved Between E. coli and Rattus (Coli/Rat)	Residue Conserved Between Bacteria and Eukaryotes?
C839	U	G	U/G	No
C2098	G	A	G/A	No
G2099	A		A/G	No
A2100				Yes
G2102				Yes
A2103				Yes
A2538				Yes
U2539				Yes
G2540				Yes
U2541				Yes
C2644	U	U		Yes
G2646	C	U	C/U	No

Residues were determined by a 5.8 angstrom distance measurement between the atoms of spiramycin and the 50S ribosome using
the program SPOCK[a]. Conserved residues were determined by comparison between the proposed secondary structures of *H. marismortui*[b], *E. coli*[b], and *Rattus norvegicus*[c].

a) Christopher, Jon A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A&M University, College Station, TX.
b) Comparative RNA Web Site URL:http://www.rna.icmb.utexas.edu/
c) Wool, I.G. (1985). "Studies of the Structure of Eukaryotic (Mammalian) Ribosomes" in Structure, Function, and Genetics of Ribosomes, *supra*, pp. 391-411.

The skilled artisan, when in possession of the foregoing or other exemplary target sites, may use the process of rational drug design to identify molecules that potentially bind to one or more of the target sites and/or inhibit ribosomal activity. Furthermore, by taking into account which of the residues that define the target site are conserved between pathogens but not conserved between host species, the skilled artisan can design new species-specific protein synthesis inhibitors. It is apparent that the skilled artisan can take advantage of the regions that are not conserved between *E. coli* and rat to provide target regions for rational drug design. By way of example, Figure 29 shows certain regions of the polypeptide exit tunnel that are conserved between *E. coli* and rat (denoted in red) and regions of the polypeptide exit tunnel that are not conserved between *E. coli* and rat (denoted in blue). Figures 29(A) and 29(B) provide enlarged views of a large ribosomal subunit when cut in half along the polypeptide exit tunnel. Figure 29(C) is provided to orient the reader to the view in Figure 29(A) relative to the large ribosomal subunit. Figure 29(D) is provided to orient the reader to the view in Figure 29B

relative to the large ribosomal subunit. In addition, the skilled artisan when in possession of mutations that prevent or reduce antibiotic activity (*i.e.*, are related to antibiotic resistance) can use this information to model the relevant antibiotic binding product which can then be used as a basis for rational drug design to identify small molecules that overcome drug resistance. It is contemplated that a variety of computer modeling procedures, for example, homology modeling protocols, can be used to provide a model of a drug resistance target site by implementing site directed mutagenesis of nucleotides and/or amino acids and then using the appropriate energy minimization and refinement protocols.

c. **Identification of Candidate Molecules.**

It is contemplated that candidate molecules that inhibit protein biosynthesis can be designed entirely *de novo* or may be based upon a pre-existing protein biosynthesis inhibitor. Either of these approaches can be facilitated by computationally screening databases and libraries of small molecules for chemical entities, agents, ligands, or compounds that can bind in whole, or in part, to ribosomes and ribosomal subunits, more preferably to large ribosomal subunits, and even more preferably to 50S ribosomal subunits. In this screening, the quality of fit of such entities or compounds to the binding site or sites may be judged either by shape complementarity or by estimated interaction energy (Meng *et al.* (1992) *J. Coma. Chem.* 13: 505-524).

The design of molecules that bind to or inhibit the functional activity of ribosomes or ribosomal subunits according to this invention generally involves consideration of two factors. First, the molecule must be capable of physically and structurally associating with the large ribosomal subunit. Non-covalent molecular interactions important in the association of ribosomes and ribosomal subunits with the molecule, include hydrogen bonding, van der Waals and hydrophobic interactions. Second, the molecule must be able to assume a conformation that allows it to associate with the ribosomes or ribosomal subunits, more preferably with the large ribosomal subunits, and even more preferably with the 50S ribosomal subunit. Although certain portions of the molecule may not directly participate in this association with a ribosome or ribosomal subunits those portions may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on binding affinities, therapeutic efficacy, drug-like qualities, and potency. Such conformational requirements include the overall three-dimensional

structure and orientation of the chemical entity or molecule in relation to all or a portion of the active site or other region of the ribosomes or ribosomal subunits, or the spacing between functional groups of a molecule comprising several chemical entities that directly interact with the ribosomes or ribosomal subunits, more preferably with the large ribosomal subunits, and even more preferably with the 50S ribosomal subunit.

The potential, predicted, inhibitory or binding effect of a molecule on ribosomes and ribosomal subunits may be analyzed prior to its actual synthesis and testing by the use of computer modeling techniques. If the theoretical structure of the given molecule suggests insufficient interaction and association between it and ribosomes or ribosomal subunits, synthesis and testing of the molecule is obviated. However, if computer modeling indicates a strong interaction, the molecule may then be synthesized and tested for its ability to interact with the ribosomes or ribosomal subunits and inhibit protein synthesis. In this manner, synthesis of inoperative molecules may be avoided. In some cases, inactive molecules are synthesized predicted on modeling and then tested to develop a SAR (structure-activity relationship) for molecules interacting with a specific region of the ribosome or ribosomal subunit, more preferably of the large ribosomal subunit, and even more preferably of the 50S ribosomal subunit. As used herein, the term "SAR", shall collectively refer to the structure-activity/structure property relationships pertaining to the relationship(s) between a compound's activity/properties and its chemical structure.

d. <u>De Novo Design</u>.

One skilled in the art may use one of several methods to identify chemical moieties or entities, compounds, or other agents for their ability to associate with a preselected target site within a ribosomes or ribosomal subunit. This process may begin by visual inspection or computer assisted modeling of, for example, the target site on the computer screen based on the atomic co-ordinates of the 50S ribosomal subunit and/or its complexes with other analogues and antibiotics, deposited in the RCSB Protein Data Bank with accession numbers PDB ID: 1FFK, 1JJ2, 1FFZ, or 1FG0, and/or listed in a table contained on Disk No. 1, 2 or 3 of 3. In one embodiment, compound design uses computer modeling programs which calculate how different molecules interact with the various sites of the ribosome, ribosomal subunit, or a fragment thereof. Selected chemical moieties or entities, compounds, or agents may then be positioned in

a variety of orientations, or docked, within at least a portion of the target site of a ribosome or ribosomal subunit, more preferably of a large ribosomal subunit, and even more preferably of a 50S ribosomal subunit. Databases of chemical structures are available from, for example, Cambridge Crystallographic Data Center (Cambridge, U.K.) and Chemical Abstracts Service (Columbus, OH). Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the process of selecting chemical entities. These include, but are not limited to:

(1) GRID (Goodford, P. J., "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules" (1985) *J. Med. Chem.* 28, 849-857). Software such as GRID, a program that determines probable interaction sites between probes with various functional group characteristics and the macromolecular surface, can be used to analyze the surface sites to determine structures of similar inhibiting proteins or molecules. The GRID calculations, with suitable inhibiting groups on molecules (*e.g.*, protonated primary amines) as the probe, are used to identify potential hotspots around accessible positions at suitable energy contour levels. GRID is available from Oxford University, Oxford, UK.

(2) MCSS (Miranker, A. and M. Karplus (1991) "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method." *Proteins: Structure, Function and Genetics* 11: 29-34). MCSS is available from Molecular Simulations, Burlington, MA.

(3) AUTODOCK (Goodsell, D. S. and A. J. Olsen (1990) "Automated Docking of Substrates to Proteins by Simulated Annealing" *Proteins: Structure, Function, and Genetics* 8: 195-202). AUTODOCK is available from Scripps Research Institute, La Jolla, CA.

(4) DOCK (Kuntz, I. D. *et al.* (1982) "A Geometric Approach to Macromolecule-Ligand Interactions" *J. Mol. Biol.* 161: 269-288). The program DOCK may be

used to analyze an active site or ligand binding site and suggest ligands with complementary steric properties. DOCK is available from University of California, San Francisco, CA.

(5) ALADDIN (Van Drie *et al.* (1989) "ALADDIN: An Integrated Tool of Computer Assisted Molecular Design and Pharmacophore Recognition From Geometric, Steric and Substructure Searching of Three-Dimensional Structures" *J. Comp-Aided Mol. Des.* 3: 225).

(6) CLIX (Davie and Lawrence (1992) "CLIX: A Search Algorithm for Funding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure" *Proteins* 12: 31-41).

(7) GROUPBUILD (Rotstein and Murcko (1993) "GroupBuild: A Fragment-Based Method for De Novo Drug Design" *J. Med. Chem* 36: 1700).

(8) GROW (Moon and Howe (1991) "Computer Design of Bioactive Molecules: A Method for Receptor-Based De Novo Ligand Design" *Proteins* 11: 314).

Once suitable chemical moieties or entities, compounds, or agents have been selected, they can be assembled into a single molecule. Assembly may proceed by visual inspection and/or computer modeling and computational analysis of the spatial relationship of the chemical moieties or entities, compounds or agents with respect to one another in three-dimensional space. This could then be followed by model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities, compounds, or agents include but are not limited to:

(1) CAVEAT (Bartlett, P. A. *et al.* (1989) "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules". In molecular Recognition in Chemical and Biological Problems", Special Pub., *Royal Chem. Soc.* 78: 82-196) and (Bacon *et al.* (1992) *J. Mol. Biol.* 225: 849-858). CAVEAT uses databases of cyclic compounds which can act as "spacers" to connect any number of chemical fragments already positioned in the active site. This allows one skilled in the art to quickly generate hundreds of possible ways to connect the

fragments already known or suspected to be necessary for tight binding. CAVEAT is available from the University of California, Berkeley, CA.

(2) 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, (CA). This area is reviewed in Martin, Y. C., (1992) "3D Database Searching in Drug Design", *J. Med. Chem.* 35: 2145-2154.

(3) HOOK (available from Molecular Simulations, Burlington, MA.).

Instead of proceeding to build a molecule of interest in a step-wise fashion one chemical entity at a time as described above, the molecule of interest may be designed as a whole using either an empty active site or optionally including some portion or portions of a known inhibitor or inhibitors. Software that implements these methods include:

(1) LUDI (Bohm, H.-J. (1992) "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", *J. ComR. Aid. Molec. Design* 6: 61-78). The program LUDI can determine a list of interaction sites into which to place both hydrogen bonding and hydrophobic fragments. LUDI then uses a library of approximately 600 linkers to connect up to four different interaction sites into fragments. Then smaller "bridging" groups such as $-CH_2-$ and -COO- are used to connect these fragments. For example, for the enzyme DHFR, the placements of key functional groups in the well-known inhibitor methotrexate were reproduced by LUDI. See also, Rotstein and Murcko, (1992) *J. Med. Chem.* 36:1700-1710. LUDI is available from Biosym Technologies, San Diego, CA.

(2) LEGEND (Nishibata, Y. and A. Itai (1991) *Tetrahedron* 47, 8985). LEGEND is available from Molecular Simulations, Burlington, MA.

(3) LeapFrog (available from Tripos Associates, St. Louis, MO.).

(4) Aladdin (available from Daylight Chemical Information Systems, Irvine, CA)

Other molecular modeling techniques may also be employed in accordance with this invention. See, *e.g.,* Cohen, N. C. *et al.* (1990) "Molecular Modeling Software and Methods for Medicinal Chemistry, *J. Med. Chem.* 33: 883-894. See also, Navia, M. A. and M. A. Murcko (1992) "The Use of Structural Information in Drug Design", *Current Opinions in Structural Biology* 2: 202-210; and Jorgensen (1998) "BOSS- Biochemical and Organic Simulation

System" in the Encyclopedia of Computational Chemistry (P.V.R. Schleyer, ed.) Wiley & Sonstra., Athens, U.S.A. 5: 3281-3285).

It is contemplated that during modeling, it may be possible to introduce into the molecule of interest, chemical moieties that may be beneficial for a molecule that is to be administered as a pharmaceutical. For example, it may be possible to introduce into or omit from the molecule of interest, chemical moieties that may not directly affect binding of the molecule to the target area but which contribute, for example, to the overall solubility of the molecule in a pharmaceutically acceptable carrier, the bioavailability of the molecule and/or the toxicity of the molecule. Considerations and methods for optimizing the pharmacology of the molecules of interest can be found, for example, in "Goodman and Gilman's The Pharmacological Basis of Therapeutics" Eighth Edition (Goodman Gilman, Rall, Nies, & Taylor (eds.)). Pergaman Press (1985); Jorgensen & Duffy (2000) *Bioorg. Med. Chem. Lett.* 10: 1155-1158.

Furthermore, the computer program "Qik Prop" can be used to provide rapid predictions for physically significant descriptions and pharmaceutically-relevant properties of an organic molecule of interest. A 'Rule of Five' probability scheme can be used to estimate oral absorption of the newly synthesized compounds (Lipinski *et al.* (1997) *Adv. Drug Deliv. Rev.* 23:3).

Programs suitable for pharmacophore selection and design include:

(1) DISCO (Abbot Laboratories, Abbot Park, Ill.).

(2) Catalyst (Bio-CAD Corp., Mountain View, CA).

(3) Chem DBS-3D (Chemical Design Ltd., Oxford, U.K.).

Furthermore, the skilled artisan may use the information available on how to design suitable therapeutically active and pharmaceutically useful compounds, and use this information in the design of new protein synthesis inhibitors of the invention. See, for example, Lipinski *et al.* (1997) *Ad. Drug Deliv. Reviews* 23: 3-25; Van de Waterbeemd *et al.* (1996) *Quantitative Structure-Activity Relationships* 15: 480-490; and Cruciani *et al* (2000), *Theochem-J. Mol. Struct.* 503: 17-30.

The entry of the co-ordinates of the ribosome's or ribosomal subunit's proteins and RNAs into the computer programs discussed above results in the calculation of most probable structure of the macromolecule, including overall atomic co-ordinates of a ribosome, ribosomal

subunit or a fragment thereof. These structures can be combined and refined by additional calculations using such programs to determine the probable or actual three-dimensional structure of the ribosome, ribosomal subunit or a fragment thereof, including potential or actual active or binding sites of ligands.

e. **Modification of Existing Molecules.**

Instead of designing molecules of interest entirely *de novo* it is contemplated that pre-existing molecules or proteins thereof may be used as a starting point for the design of a new candidate. It is contemplated that many of the approaches useful for designing molecules *de novo* may also be useful for modifying existing molecules.

It is contemplated that knowledge of the spatial relationship between a protein biosynthesis inhibitor, for example, an antibiotic, and its respective binding site within a ribosome permits the design of modified inhibitors that may have better binding properties, for example, higher binding affinity and/or specificity, relative to the molecule from which it was derived. Alternatively, knowledge of inhibitor contact sites within a ribosome permits the synthesis of a new molecule that contain, for example, a portion of a first molecule that binds to the contact site and another portion that contributes additional functionality.

It is contemplated that a variety of modified molecules (for example, modified antibiotics) may be designed using the atomic co-ordinates provided herein. For example, it is contemplated that by knowing the spatial relationship of one or more of antibiotics relative to the large ribosomal subunit it is possible to generate new antibiotic-based molecules. The atomic co-ordinates of each antibiotic relative to the large ribosomal subunit provides information on what portions of the ribosome or ribosomal subunit and the antibiotic contact one another. Accordingly, from this information the skilled artisan may not only identify contact locations within the ribosome that can be used for *de novo* drug design, as discussed above, but also may identify portions of an antibiotic that can act as a ribosome binding domain.

Based on the information provided herein, the skilled artisan may readily identify and produce hybrid antibiotics that comprise a ribosome binding domain of a first antibiotic and a ribosome binding domain of a second, different antibiotic. The resulting hybrid antibiotics preferably bind to each of respective contact locations within the ribosomal subunit simultaneously. The atomic co-ordinates provided herein permit the skilled artisan to identify

candidate antibiotics that may be used as templates in the synthesis of a hybrid, and also provide steric information necessary to produce linking chemistries such that each ribosome binding domain is properly orientated relative to its respective contact site. As a result, it is contemplated that the skilled artisan may produce a hybrid antibiotic that binds to a ribosome or ribosomal subunit with a higher affinity and/or have higher protein synthesis inhibitory activity than either of the individual template antibiotics used to generate the hybrid. Alternatively, the hybrid antibiotic may overcome resistance phenotypes that may have developed against either of the template antibiotics. For example, the proximity of the site occupied by the disaccharide moiety of carbomycin to the site filled by anisomycin suggests that a hybrid compound including portions of both carbomycin and anisomycin may be an effective inhibitor of protein synthesis.

Furthermore, the atomic co-ordinates provided herein permit the skilled artisan to use the information pertaining to identify a ribosome binding domain and to design other types of protein synthesis inhibitors. For example, with an understanding of the ribosome contact region and the surrounding environment, the skilled artisan can provide novel molecules, a portion of which is based upon the antibiotic binding region (binding domain) and another portion of which (effector domain) can be designed as a novel space filling domain that sterically inhibits or disrupts protein biosynthesis within the ribosome or secretion through the polypeptide exit tunnel. For example, the skilled artisan may combine the ribosome binding region of the antibiotic, tylosin, which binds to one side of the polypeptide exit tunnel close to the peptidyl transferase site, with a novel chemical moiety bulky enough to block the polypeptide exit tunnel. However, it is contemplated that the skilled artisan may take advantage of one or more of the many of the antibiotic contact regions disclosed herein to design entirely new binding and effector domains.

Furthermore, the present invention permits the skilled artisan to design molecules, for example, selective protein synthesis inhibitors that are tailored to be more potent with respect to ribosomes of a target organism, for example, a pathogen such a microbe, and less potent, *i.e.,* less toxic, to ribosomes of a non target organism, for example, host organism such as a human. Also, the invention permits the skilled artisan to use the atomic co-ordinates and structures of the large ribosomal subunit and its complexes with protein synthesis inhibitors to design modifications to starting compounds, such as an antibiotic, that will bind more tightly to a target

ribosome (*e.g.*, the 50S ribosomal subunit of bacteria) and less tightly to a non-targeted ribosome (*e.g.*, human 60S ribosomal subunit).

The structure of a complex between the large ribosomal subunit and the starting compound (*e.g.*, tylosin or another protein synthesis inhibitor) can also be used to guide the modification of that compound to produce new compounds that have other desirable properties for the applicable industrial and other uses (*e.g.*, as pharmaceuticals, herbicides or insecticides), such as chemical stability, solubility or membrane permeability.

A variety of antibiotics bind the large ribosomal subunit and disrupt protein synthesis and include members of antibiotic families which include, for example, chloramphenicols, macrolides, lincosamides, streptogramins, althiomycins, oxazolinones, nucleotide analogs, thiostreptons, peptides, glutarimides, and trichothecenes.

Members of the chloramphenicol family include, for example, Chloramphenicol and Iodoamphenicol. Members of the macrolide family include, for example, Biaxin (Clarithromycin), Zithromax (Azithromycins), Ketek (Telithromycin; ketolide), ABT-773, Tylosin, Spiramycin I, Spiramycin II, Spiramycin III, Erythromycin A, Carbomycin A, Telithromycin, Methymycin, Narbomycin, Lankamycin, Oleandomycin, Megalomycin, Chalcomycin, Niddamycin, Leucomycin, Angolamycin, and Relomycin. Members of the licosamide family include, for example, Clindamycin and Lincomycin. Members of the streptogramin family include, for example, Streptogramin A, Streptogramin B, Ostreogrycin G, Synercid, Virginiamycin S1, Virginiamycin S2, Virginiamycin S3, Virginiamycin S4, Vernamycin B, Vernamycin C, Patricin A, and Patricin B. A member of the althiomycin family, includes, for example, Althiomycin. A member of the oxazolidine family, includes, for example, Linezolid. Members of the family of nucleotide analogs include, for example, Sparsomycin, Puromycin, Anisomycin, and Blasticidin S. Members of the thiostrepton family include, for example, Thiostrepton, Siomycin, Sporangiomycin, and Thiopeptin. Members of the peptide family include, for example, Viomycin, Capreomycin IA, Capreomycin IB, Capreomycin IIA, and Capreomycin IIB. Members of the glutarimide family include, for example, Cycloheximide, Streptovitacins, Streptimidone, Inactone, Actiphenol. Members of the trichothecene family include, for example, Trichodermin, Trichodermol, Trichodermone, Vomitoxin, T-2 toxin, Trichothecin, Nivalenol, and Verrucarin A.

Inhibitors can be diffused into or soaked with the stabilized crystals of the large ribosomal subunit as described in Example 3 to form a complex with the large ribosomal subunit for collecting X-ray diffraction data. Alternatively, the inhibitors can be co-crystallized with the large ribosomal subunit by mixing the inhibitor with the large ribosomal subunit before precipitation with high salt.

Starting with the structure of the ribosome from *H. marismortui*, the structure of the ribosome from a non-targeted organism (for example, the human 60S ribosomal subunit) can be constructed by homology modeling, *i.e.*, by changing the structure of residues at a target site of interest for the residues at the same positions in of the non-target ribosome. This is done computationally by removing the side chains from the ribosome of known structure and replacing them by the side chains of the unknown structure put in sterically plausible positions. In this way, it can be understood how the shapes of the target sites within the targeted and non-targeted ribosomes differ. This process, therefore, provides information concerning how a molecule that binds the target site can be chemically altered in order to produce molecules that will bind tightly and specifically to the targeted ribosome but will simultaneously be prevented from binding to the non-targeted ribosome. Likewise, knowledge of portions of the bound molecules that face the solvent permit introduction of other functional groups for additional pharmaceutical purposes. The process of homology structure modeling can also be used to understand the mechanisms whereby mutant ribosomes become resistant to the effects of pharmaceuticals or pesticides, such as herbicides or insecticides. Furthermore, with knowledge of the portions of the ribosomal subunit that participates in drug resistance, the skilled artisan may design new molecules that overcome the problem of drug resistance.

The use of homology structure modeling to design molecules that bind more tightly to the target ribosome than to the non-target ribosome has wide-spread applicability. The methods outlined herein can be used to control any targeted organism, for example, a pathogen, by designing molecules that inhibit large ribosomal subunits of the targeted organisms while failing to inhibit the 50S or 60S ribosomal subunit of the non-targeted organism, for example, a host, to the same extent or not at all. The molecules identified or prepared by the methods of the present invention can be used to control the targeted organisms while causing the non-targeted organism little or no adverse effects. Thus, the molecules identified or developed using the methods of the present invention can be designed so that their administration kills the target organisms or

inhibits some aspect of the biological functions of the target organisms while failing to have a similar effect on the non-targeted organism. The adverse effects of the agent on the targeted organisms may include, but are not limited to, death of the target organism; slowing growth rates; slowing or eliminating passage from one growth phase to another (*e.g.,* extending the larval growth stage); slowing or eliminating reproduction, decreasing or preventing mating, decreasing or eliminating offspring production, limiting or eliminating target organism weight gains; decreasing or eliminating feeding ability and behaviors; and disrupting cellular, tissue and/or organ functions.

The novel agents contemplated by the present invention can be useful as herbicides, pesticides *(e.g.,* insecticides, nematocides, rodenticides, etc.), miticides, or antimicrobial agents (*e.g.,* antifungals, antibacterials, antiprotozoals, etc.) to target specific organisms. For example, the novel agents can target animal and plant parasitic nematodes, prokaryotic organisms (disease causing microbes), and eukaryotic multicellular pests. Specific examples of multicellular pests include, but are not limited to, insects, fungi, bacteria, nematodes, mites and ticks, protozoan pathogens, animal-parasitic liver flukes, and the like.

Herbicides, pesticides, miticides, and antimicrobial agents that inhibit protein synthesis by interacting with ribosomes are known to the skilled artisan. A few examples are discussed below. These known agents can be modified to obtain novel agents by using computer modeling techniques and knowledge of the structure of ribosomes and ribosomal subunits and the structure of ribosome/agent and ribosomal subunit/agent complexes.

The ketolide ABT-773 binds ribosomes tighter than erythromycin in *S. pneumoniae* and is able to defeat macrolide resistance in bacteria (Capobianco *et al.* (2000) *Antimicrob Agents Chemother* 44(6), 1562-1567). The tools and methodologies of the present invention can be used to obtain erythromycin derivatives that bind the ribosomes or ribosomal subunits of target bacteria more tightly than they bind the ribosomes and ribosomal subunits of non-target animals. The target bacteria can be any infectious bacteria, particularly *S. pneumoniae*, and even more particularly erythromycin-resistant *S. pneumoniae*. The non-target animals can be any animal, particularly mammals, and even more particularly humans.

Examples of antibiotics that are inhibitors of protein synthesis include, but are not limited to, puromycin, cycloheximide, chloramphenicol, tetracycline, and streptomycin (Heldt, (1996)

Plant Biochemistry and Molecular Biology 21.2: 458-464). Puromycin, as discussed earlier, binds as an analogue of an aminoacyl-tRNA to the A-site and is added to nascent peptide chains, its weak associate with the ribosome prevents further elongation steps in prokaryotes and eukaryotes. Cycloheximide inhibits peptidyl transferase in eukaryotic ribosomes. Chloramphenicol inhibits peptidyl transferase in prokaryotic ribosomes. Tetracycline binds to the 30S subunit and inhibits the binding of aminoacyl-tRNA to prokaryotic ribosomes much more than to eukaryotic ones. Streptomycin interacts with 30S ribosomes which results in an incorrect recognition of mRNA sequences and thus inhibits initiation in prokaryotic ribosomes. U.S. Patent No. 5,801,153 discloses antibiotics against pathogens. Aminoglycosides are examples of antibacterial antibiotics that appear to inhibit protein synthesis. However, there is a limitation to their use because of their ototoxic and nephrotoxic properties. Amikacin sulfate, Framycetin sulfate, Gentamycin sulfate, Kanamycin sulfate, Neomycin sulfate, Netilmicin sulfate, Paromomycin sulfate, Sissomycin sulfate, Tobramycin, Vancomycin hydrochloride, and Viomycin sulfate are the members of the aminoglycoside family. The tools and methodologies of the present invention can be used to obtain derivatives of any antibiotic of choice so that they inhibit the protein synthesis of target organisms to a greater degree than they inhibit the protein synthesis of non-target organisms, such as humans.

Examples of targeted and non-targeted organisms include, but are not limited to, those provided in Table 18.

Table 18

Examples of Classes of Molecules which can be Identified and/or
Developed by the Methods of the Invention and Applicable Target/Non-Target Organisms.

Type of Molecule	Target Organisms	Non-Target Organisms
Herbicides	Dicotyledonous plants	Monocotyledonous plants
Herbicides	Grasses	Soybeans, potatoes, coffee
Insecticides	Flies, Mites	Honey bees
Pesticides	Ticks	Deer
Pesticides	Lice	Birds
Miticides	Parasitic mites (mange)	Dogs
Antimicrobial Agents (Antibacterials)	*Streptococcus pneumoniae*	Humans
Antimicrobial Agents (Antibacterials)	*Clostridium difficile*	*Escherichia coli*
Antimicrobial Agents (Antifungals)	*Erysiphe graminis*	Barley
Antimicrobial Agents (Antiprotozoals)	Toxoplasma gondii	Animals
Poisons (Rodentcides)	Rats	Dogs, cats, humans

It is contemplated that the tools and methodologies of the present invention can be used to obtain inhibitors of protein synthesis of target insects, such as bollworms and mosquitoes, more than they inhibit the protein synthesis of non-target insects, such as beetles of the family Coccinellidae (*e.g.*, ladybugs) and *Apis mellifera* (honey bees). Other possible target insects include, but are not limited to, insects selected from the orders *Coleoptera* (beetles), *Diptera* (flies, mosquitoes), *Hymenoptera* (wasps, ants, sawflies), *Lepidoptera* (butterflies and moths), *Mallophaga* (lice), *Homoptera* (whiteflies, aphids), *Hemiptera* (bugs), *Orthroptera* (locusts, cockroaches), *Thysanoptera* (thrips), *Dermaptera* (earwigs), *Isoptera*, *Anoplura, Siphonaptera*, and *Trichoptera* (caddis flies).

Furthermore, it is contemplated that the tools and methodologies of the present invention can be used to obtain inhibitors of protein synthesis of target plants which inhibit protein synthesis of the target plants more than they inhibit the protein synthesis of non-target plants and animals. The target plants can be any unwanted plant species, particular weeds, and even more particularly noxious weeds. Whether or not a particular plant is considered a weed will depend upon the context in which it is growing. For example, unwanted *Zea mays* (corn) plants growing in a *Glycine max* (soybean) field could be considered unwanted weeds. Examples of weeds which are likely target plants include, but are not limited to, *Allium vineale* (wild garlic), *Bromus tectorum* (downy brome), *Triticum cylindricum* (jointed goatgrass), *Amaranthus* spp. (pigsweed), *Chenopodium album* (lambsquarters), *Avena fatua* (wild oats), *B. secalinus* (cheat), *Echinochloa crus-galli* (barnyardgrass), *Alopecurus myosuroides* (blackgrass), *Setaria faberii* (giant foxtail), *Xanthium strumarium* (common cocklebur), *Ambrosia artemisiifolia* (common ragweed), and *Ipomoea* spp. (morning glories). The non-target organisms can be any plant, particularly any desirable plant, and even more particularly any crop plant. The non-target organisms can also be any animals, particularly mammals, and even more particularly humans. In one preferred embodiment, the tools and methodologies of the present invention can be used to produce protein synthesis inhibitors which kill or injure one or more noxious weed species but fail to harm non-target plants and animals.

Target bacteria of interest include, but are not limited to, *Staphylococcus aureus, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus bovis, Streptococcus pneumoniae, Moraxella catarrhalis, Neisseria gonorrhoeae, Neisseria meningitides, Bacillus anthracis, Corynebacterium diphtheriae, Listeria monocytogenes, Erysipelothrix rhusiopathiae, Clostridium perfringens, Clostridium tetani, Clostridium difficile, Eschericia coli, Proteus mirabilis, Psuedomonas aeruginosa, Klebsiella pneumoniae, Haemophilus influenzae, Haemophilus ducreyi, Yersinia pestis, Yersinia enterocolitica, Francisella tularensis, Pasteurella multocida, Vibrio cholerae, Flavobacterium meningosepticum, Pseudomonas mallei, Pseudomonas pseudomallei, Campylobacter jejuni, Campylobacter fetus, Fusobacterium nucleatum, Calymmatobacterium granulomatis, Streptobacillus moniliformis, Legionella pneumophila, Mycobacterium avium-intracellulare, Mycobacterium tuberculosis, Mycobacterium leprae, Treponema pallidum, Treponema pertenue, Borrelia burgdorferi, Borrelia recurrentis, Actinomyces isrealii, Nocardia asteroides, Ureaplasma urealyticum, Mycoplasma pneumoniae, Chlamydia psittaci, Chlamydia trachomatis, Chlamydia pnemoniae,*

Pneumocystis carinii, Coccidioides immitis, Histoplasma capsulatum, Blastomyces dermatitidis, Paracoccidioides brasiliensis, Sporothrix schenckii, Cryptococcus neoformans.

Once a candidate molecule has been designed or selected by the above methods, the affinity with which that molecule may bind to the ribosome or ribosomal subunit may be tested and optimized by computational evaluation and/or by testing biological activity after synthesizing the compound. Candidate molecules may interact with the ribosomes or ribosomal subunits in more than one conformation each of which has a similar overall binding energy. In those cases, the deformation energy of binding may be considered to be the difference between the energy of the free molecule and the average energy of the conformations observed when the molecule binds to the ribosomes or ribosomal subunits, more preferably to the large ribosomal subunits, and even more preferably to the 50S ribosomal subunits.

A molecule designed or selected as binding to a ribosome or ribosomal subunit may be further computationally optimized so that in its bound state it preferably lacks repulsive electrostatic interaction with the target region. Such non-complementary (*e.g.*, electrostatic) interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions. Specifically, the sum of all electrostatic interactions between the inhibitor and the enzyme when the inhibitor is bound to the ribosome or the ribosomal subunit, preferably make a neutral or favorable contribution to the enthalpy of binding. Weak binding compounds can also be designed by these methods so as to determine SAR.

Specific computer programs that can evaluate a compound deformation energy and electrostatic interaction are available in the art. Examples of suitable programs include: Gaussian 92, revision C (M. J. Frisch, Gaussian, Inc., Pittsburgh, PA.); AMBER, version 4.0 (P. A. Kollman, University of California at San Francisco, CA); QUANTA/CHARMM (Molecular Simulations, Inc., Burlington, MA); and Insight II/Discover (Biosysm Technologies Inc., San Diego, CA). These programs may be implemented, for instance, using a Silicon Graphics workstation, IRIS 4D/35 or IBM RISC/6000 workstation model 550. Other hardware systems and software packages are known to those skilled in the art.

Once a molecule of interest has been selected or designed, as described above, substitutions may then be made in some of its atoms or side groups in order to improve or modify its binding properties. Generally, initial substitutions are conservative, *i.e.,* the replacement group will approximate the same size, shape, hydrophobicity and charge as the original group. It should, of course, be understood that components known in the art to alter conformation should be avoided. Such substituted chemical compounds may then be analyzed

for efficiency of fit to the ribosome or ribosomal subunit by the same computer methods described in detail, above.

In addition, the actual ribosome-related ligands, complexes or mimetics may be crystallized and analyzed using X-ray diffraction. The diffraction pattern co-ordinates are similarly used to calculate the three-dimensional interaction of a ligand and the ribosome, ribosomal subunit, or a mimetic, in order to confirm that the ligand binds to, or changes the conformation of, a particular site on the ribosome or ribosomal subunit, or where the mimetic has a similar three-dimensional structure to that of a ribosome, ribosomal subunit or a fragment thereof.

3. Synthesis of Lead Molecules

A lead molecule of the present invention can be, but is not limited to, at least one selected from a lipid, nucleic acid, peptide, small organic or inorganic molecule, chemical compound, element, saccharide, isotope, carbohydrate, imaging agent, lipoprotein, glycoprotein, enzyme, analytical probe, and an antibody or fragment thereof, any combination of any of the foregoing, and any chemical modification or variant of any of the foregoing. In addition, a lead molecule may optionally comprise a detectable label. Such labels include, but are not limited to, enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. Well known methods may be used for attaching such a detectable label to a lead molecule.

Methods useful for synthesizing lead molecules such as lipids, nucleic acids, peptides, small organic or inorganic molecules, chemical compounds, elements, saccharides, isotopes, carbohydrates, imaging agents, lipoproteins, glycoproteins, enzymes, analytical probes, antibodies, and antibody fragments are well known in the art. Such methods include the traditional approach of synthesizing one such lead molecule, such as a single defined peptide, at a time, as well as combined synthesis of multiple lead molecules in a one or more containers. Such multiple lead molecules may include one or more variants of a previously identified lead molecule. Methods for combined synthesis of multiple lead molecules are particularly useful in preparing combinatorial libraries, which may be used in screening techniques known in the art.

By way of example, it is well known in the art that multiple peptides and oligonucleotides may be simultaneously synthesized. Lead molecules that are small peptides up to 50 amino acids in length, may be synthesized using standard solid-phase peptide synthesis procedures, for example, procedures similar to those described in Merrifield (1963) *J. Am. Chem. Soc.,* 85: 2149. For example, during synthesis, N-α-protected amino acids having protected side chains are added stepwise to a growing polypeptide chain linked by its C-terminal end to an insoluble polymeric support, *e.g.,* polystyrene beads. The peptides are synthesized by linking an amino group of an N-α-deprotected amino acid to an α-carboxy group of an N- α-protected amino acid that has been activated by reacting it with a reagent such as dicyclohexylcarbodiimide. The attachment of a free amino group to the activated carboxyl leads to peptide bond formation. The most commonly used N-α-protecting groups include Boc which is acid labile and Fmoc which is base labile.

Briefly, the C-terminal N-α-protected amino acid is first attached to the polystyrene beads. Then, the N-α-protecting group is removed. The deprotected α-amino group is coupled to the activated α-carboxylate group of the next N-α-protected amino acid. The process is repeated until the desired peptide is synthesized. The resulting peptides are cleaved from the insoluble polymer support and the amino acid side chains deprotected. Longer peptides, for example greater than about 50 amino acids in length, typically are derived by condensation of protected peptide fragments. Details of appropriate chemistries, resins, protecting groups, protected amino acids and reagents are well known in the art and so are not discussed in detail herein. See for example, Atherton *et al.* (1963) Solid Phase Peptide Synthesis: A Practical Approach (IRL Press), and Bodanszky (1993) Peptide Chemistry, A Practical Textbook, 2nd Ed. Springer-Verlag, and Fields *et al.* (1990) *Int. J. Peptide Protein Res. 35*:161-214.

Purification of the resulting peptide is accomplished using conventional procedures, such as preparative HPLC, *e.g.,* gel permeation, partition and/or ion exchange chromatography. The choice of appropriate matrices and buffers are well known in the art and so are not described in detail herein.

It is contemplated that a synthetic peptide in accordance with the invention may comprise naturally occurring amino acids, unnatural amino acids, and/or amino acids having specific characteristics, such as, for example, amino acids that are positively charged, negatively charged,

hydrophobic, hydrophilic, or aromatic. As used herein, the term "naturally occurring amino acids" refers to the L-isomers of amino acids normally found in proteins. The predominant naturally occurring amino acids are glycine, alanine, valine, leucine, isoleucine, serine, methionine, threonine, phenylalanine, tyrosine, tryptophan, cysteine, proline, histidine, aspartic acid, asparagine, glutamic acid, glutamine, arginine, and lysine. Unless specifically indicated, all amino acids are referred to in this application are in the L-form. Furthermore, as used herein, the term "unnatural amino acids" refers to amino acids that are not naturally found in proteins. For example, selenomethionine.

Amino acids that are "positively charged" include any amino acid having a positively charged side chain under normal physiological conditions. Examples of positively charged naturally occurring amino acids include, for example, arginine, lysine, and histidine. Conversely, amino acids that are "negatively charged" include any amino acid having a negatively charged side chains under normal physiological conditions. Examples of negatively charged naturally occurring amino acids include, for example, aspartic acid and glutamic acid.

As used herein, the term "hydrophobic amino acid" includes any amino acids having an uncharged, nonpolar side chain that is relatively insoluble in water. Examples of naturally occurring hydrophobic amino acids include, for example, alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine. In addition, as used herein, the term "hydrophilic amino acid" refers to any amino acids having an uncharged, polar side chain that is relatively soluble in water. Examples of naturally occurring hydrophilic amino acids include, for example, serine, threonine, tyrosine, asparagine, glutamine and cysteine.

Finally, as used herein, the term "aromatic" refers to amino acid residues which side chains have delocalized conjugated system. Examples of aromatic residues include, for example, phenylalanine, tryptophan, and tyrosine.

With regard to the production of non-peptide small organic molecules which act as a ligand in the present invention, these molecules can be synthesized using standard organic chemistries well known and thoroughly documented in the patent and other literatures.

Many of the known methods useful in synthesizing lead of the present invention may be automated, or may otherwise be practiced on a commercial scale. As such, once a lead molecule

has been identified as having commercial potential, mass quantities of that molecule may easily be produced.

4. Characterization of Molecules

Molecules designed, selected and/or optimized by methods described above, once produced, may be characterized using a variety of assays known to those skilled in the art to determine whether the compounds have biological activity. For example, the molecules may be characterized by conventional assays, including but not limited to those assays described below, to determine whether they have a predicted activity, binding activity and/or binding specificity.

Furthermore, high-throughput screening may be used to speed up analysis using such assays. As a result, it may be possible to rapidly screen new molecules for their ability to interact with a ribosome or ribosomal subunit using the tools and methods of the present invention. General methodologies for performing high-throughput screening are described, for example, in Devlin, (1998), High Throughput Screening, Marcel Dekker; and U.S. Patent No. 5,763,263. High-throughput assays can use one or more different assay techniques including, but not limited to, those described below.

(1) *Surface Binding Studies*. A variety of binding assays may be useful in screening new molecules for their binding activity. One approach includes surface plasmon resonance (SPR) which can be used to evaluate the binding properties molecules of interest with respect to a ribosome, ribosomal subunit or a fragment thereof.

SPR methodologies measure the interaction between two or more macromolecules in real-time through the generation of a quantum-mechanical surface plasmon. One device, (BlAcore Biosensor RTM from Pharmacia Biosensor, Piscatawy, N.J.) provides a focused beam of polychromatic light to the interface between a gold film (provided as a disposable biosensor "chip") and a buffer compartment that can be regulated by the user. A 100 nm thick "hydrogel" composed of carboxylated dextran which provides a matrix for the covalent immobilization of analytes of interest is attached to the gold film. When the focused light interacts with the free electron cloud of the gold film, plasmon resonance is enhanced. The resulting reflected light is spectrally depleted in wavelengths that optimally evolved the resonance. By separating the reflected polychromatic light into its component wavelengths (by means of a prism), and

determining the frequencies which are depleted, the BIAcore establishes an optical interface which accurately reports the behavior of the generated surface plasmon resonance. When designed as above, the plasmon resonance (and thus the depletion spectrum) is sensitive to mass in the evanescent field (which corresponds roughly to the thickness of the hydrogel). If one component of an interacting pair is immobilized to the hydrogel, and the interacting partner is provided through the buffer compartment, the interaction between the two components can be measured in real time based on the accumulation of mass in the evanescent field and its corresponding effects of the plasmon resonance as measured by the depletion spectrum. This system permits rapid and sensitive real-time measurement of the molecular interactions without the need to label either component.

(2) *Immunodiagnostics and Immunoassays*. These are a group of techniques that can be used for the measurement of specific biochemical substances, commonly at low concentrations in complex mixtures such as biological fluids, that depend upon the specificity and high affinity shown by suitably prepared and selected antibodies for their complementary antigens. A substance to be measured must, of necessity, be antigenic - either an immunogenic macromolecule or a haptenic small molecule. To each sample a known, limited amount of specific antibody is added and the fraction of the antigen combining with it, often expressed as the bound:free ratio, is estimated, using as indicator a form of the antigen labeled with radioisotope (radioimmunoassay), fluorescent molecule (fluoroimmunoassay), stable free radical (spin immunoassay), enzyme (enzyme immunoassay), or other readily distinguishable label.

Antibodies can be labeled in various ways, including: enzyme-linked immunosorbent assay (ELISA); radioimmuno assay (RIA); fluorescent immunoassay (FIA); chemiluminescent immunoassay (CLIA); and labeling the antibody with colloidal gold particles (immunogold).

Common assay formats include the sandwich assay, competitive or competition assay, latex agglutination assay, homogeneous assay, microtitre plate format and the microparticle-based assay.

(3) *Enzyme-linked immunosorbent assay (ELISA)*. ELISA is an immunochemical technique that avoids the hazards of radiochemicals and the expense of fluorescence detection systems. Instead, the assay uses enzymes as indicators. ELISA is a form of quantitative immunoassay based on the use of antibodies (or antigens) that are linked to an insoluble carrier

surface, which is then used to "capture" the relevant antigen (or antibody) in the test solution. The antigen-antibody complex is then detected by measuring the activity of an appropriate enzyme that had previously been covalently attached to the antigen (or antibody).

General methods and compositions for practicing ELISA are described, for example, in Crowther (1995) ELISA - Theory and Practice (Methods in Molecular Biology), Humana Press; Challacombe and Kemeny, (1998) ELISA and Other Solid Phase Immunoassays - Theoretical and Practical Aspects, John Wiley; Kemeny, (1991) A Practical Guide to ELISA, Pergamon Press; Ishikawa, (1991) Ultrasensitive and Rapid Enzyme Immunoassay (Laboratory Techniques in Biochemistry and Molecular Biology) Elsevier.

(4) *Colorimetric Assays.* Colorimetry is any method of quantitative chemical analysis in which the concentration or amount of a compound is determined by comparing the color produced by the reaction of a reagent with both standard and test amounts of the compound, often using a colorimeter. A colorimeter is a device for measuring color intensity or differences in color intensity, either visually or photoelectrically.

Standard colorimetric assays of beta-galactosidase enzymatic activity are well known to those skilled in the art (see, for example, Norton *et al.* (1985) *Mol. Cell. Biol.* 5: 281-290). A colorimetric assay can be performed on whole cell lysates using O-nitrophenyl-β-D-galactopyranoside (ONPG, Sigma) as the substrate in a standard colorimetric beta-galactosidase assay (Sambrook *et al.* (1989) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory Press). Automated colorimetric assays are also available for the detection of β-galactosidase activity, as described in U.S. Patent No. 5,733,720.

(5) *Immunofluorescence Assays.* Immunofluorescence or immunofluorescence microscopy is a technique in which an antigen or antibody is made fluorescent by conjugation to a fluorescent dye and then allowed to react with the complementary antibody or antigen in a tissue section or smear. The location of the antigen or antibody can then be determined by observing the fluorescence by microscopy under ultraviolet light.

A general description of immunofluorescent techniques appears for example, in Knapp *et al.* (1978) Immunofluorescence and Related Staining Techniques, Elsevier; Allan, (1999) Protein Localization by Fluorescent Microscopy - A Practical Approach (The Practical Approach

Series) Oxford University Press; Caul, (1993) Immunofluorescence Antigen Detection Techniques in Diagnostic Microbiology, Cambridge University Press. For detailed explanations of immunofluorescent techniques applicable to the present invention, see, for example, U.S. Patent No. 5,912,176; U.S. Patent No. 5,869,264; U.S. Patent No. 5,866,319; and U.S. Patent No. 5,861,259.

(6) *Fluorescence Polarization.* Fluorescence polarization (FP) is a measurement technique that can readily be applied to protein-protein and protein-ligand interactions in order to derive IC_{50}s and Kds of the association reaction between two molecules. In this technique one of the molecules of interest is conjugated with a fluorophore. This is generally the smaller molecule in the system (in this case, the molecule of interest). The sample mixture, containing both the ligand-probe conjugate and the ribosome, ribosomal subunit or fragment thereof, is excited with vertically polarized light. Light is absorbed by the probe fluorophores, and re-emitted a short time later. The degree of polarization of the emitted light is measured. Polarization of the emitted light is dependent on several factors, but most importantly on viscosity of the solution and on the apparent molecular weight of the fluorophore. With proper controls, changes in the degree of polarization of the emitted light depends only on changes in the apparent molecular weight of the fluorophore, which in-turn depends on whether the probe-ligand conjugate is free in solution, or is bound to a receptor. Binding assays based on FP have a number of important advantages, including the measurement of IC_{50}s and Kds under true homogenous equilibrium conditions, speed of analysis and amenity to automation, and ability to screen in cloudy suspensions and colored solutions.

(7) *Protein Synthesis.* It is contemplated that, in addition to characterization by the foregoing biochemical assays, the molecule of interest may also be characterized as a modulator (for example, an inducer of protein synthesis or an inhibitor of protein synthesis) of the functional activity of the ribosome or ribosomal subunit.

Inhibitors of protein synthesis may be assayed on the cellular level. For example, molecules of interest can be assayed for inhibitory action against organisms, for example, micro-organism, by growing the micro-organism of interest in media either containing or lacking the molecule of interest. Growth inhibition may be indicative that the molecule may be acting as a protein synthesis inhibitor.

Furthermore, more specific protein synthesis inhibition assays may be performed by administering the compound to a whole organism, tissue, organ, organelle, cell, a cellular or subcellular extract, or a purified ribosome preparation and observing its pharmacological and inhibitory properties by determining, for example, its inhibition constant (IC_{50}) for inhibiting protein synthesis. Incorporation of ^{3}H leucine or ^{35}S methionine, or similar experiments can be performed to investigate protein synthesis activity.

A change in the amount or the rate of protein synthesis in the cell in the presence of a molecule of interest indicates that the molecule is an inducer of protein synthesis. A decrease in the rate or the amount of protein synthesis indicates that the molecule is a inhibitor of protein synthesis.

H. Drug Formulation and Administration

It is contemplated that once identified, the active molecules of the invention may be incorporated into any suitable carrier prior to use. More specifically, the dose of active molecule, mode of administration and use of suitable carrier will depend upon the target and non-target organism of interest.

It is contemplated that with regard to mammalian recipients, the compounds of interest may be administered by any conventional approach known and/or used in the art. Thus, as appropriate, administration can be oral or parenteral, including intravenous and intraperitoneal routes of administration. In addition, administration can be by periodic injections of a bolus, or can be made more continuous by intravenous or intraperitoneal administration from a reservoir which is external (*e.g.,* an intrvenous bag). In certain embodiments, the compounds of the invention can be therapeutic-grade. That is, certain embodiments comply with standards of purity and quality control required for administration to humans. Veterinary applications are also within the intended meaning as used herein.

The formulations, both for veterinary and for human medical use, of the drugs according to the present invention typically include such drugs in association with a pharmaceutically acceptable carrier therefore and optionally other therapeutic ingredient(s). The carrier(s) should be "acceptable" in the sense of being compatible with the other ingredients of the formulations and not deleterious to the recipient thereof. Pharmaceutically acceptable carriers, in this regard,

are intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds (identified or designed according to the invention and/or known in the art) also can be incorporated into the compositions. The formulations may conveniently be presented in dosage unit form and may be prepared by any of the methods well known in the art of pharmacy/microbiology. In general, some formulations are prepared by bringing the drug into association with a liquid carrier or a finely divided solid carrier or both, and then, if necessary, shaping the product into the desired formulation.

A pharmaceutical composition of the invention should be formulated to be compatible with its intended route of administration. Examples of routes of administration include oral or parenteral, *e.g.,* intravenous, intradermal, inhalation, transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide.

Useful solutions for oral or parenteral administration can be prepared by any of the methods well known in the pharmaceutical art, described, for example, in Remington's Pharmaceutical Sciences, (Gennaro, A., ed.), Mack Pub., (1990). Formulations for parenteral administration can also include glycocholate for buccal administration, methoxysalicylate for rectal administration, or cutric acid for vaginal administration. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic. Suppositories for rectal administration also can be prepared by mixing the drug with a non-irritating excipient such as cocoa butter, other glycerides, or other compositions which are solid

at room temperature and liquid at body temperatures. Formulations also can include, for example, polyalkylene glycols such as polyethylene glycol, oils of vegetable origin, hydrogenated naphthalenes, and the like. Formulations for direct administration can include glycerol and other compositions of high viscosity. Other potentially useful parenteral carriers for these drugs include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes. Formulations for inhalation administration can contain as excipients, for example, lactose, or can be aqueous solutions containing, for example, polyoxyethylene-9-lauryl ether, glycocholate and deoxycholate, or oily solutions for administration in the form of nasal drops, or as a gel to be applied intranasally. Retention enemas also can be used for rectal delivery.

Formulations of the present invention suitable for oral administration may be in the form of discrete units such as capsules, gelatin capsules, sachets, tablets, troches, or lozenges, each containing a predetermined amount of the drug; in the form of a powder or granules; in the form of a solution or a suspension in an aqueous liquid or non-aqueous liquid; or in the form of an oil-in-water emulsion or a water-in-oil emulsion. The drug may also be administered in the form of a bolus, electuary or paste. A tablet may be made by compressing or moulding the drug optionally with one or more accessory ingredients. Compressed tablets may be prepared by compressing, in a suitable machine, the drug in a free-flowing form such as a powder or granules, optionally mixed by a binder, lubricant, inert diluent, surface active or dispersing agent. Moulded tablets may be made by moulding, in a suitable machine, a mixture of the powdered drug and suitable carrier moistened with an inert liquid diluent.

Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients. Oral compositions prepared using a fluid carrier for use as a mouthwash include the compound in the fluid carrier and are applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose; a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes;

a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELTM (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition should be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and should be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation include vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Formulations suitable for intra-articular administration may be in the form of a sterile aqueous preparation of the drug which may be in microcrystalline form, for example, in the form

of an aqueous microcrystalline suspension. Liposomal formulations or biodegradable polymer systems may also be used to present the drug for both intra-articular and ophthalmic administration.

Formulations suitable for topical administration, including eye treatment, include liquid or semi-liquid preparations such as liniments, lotions, gels, applicants, oil-in-water or water-in-oil emulsions such as creams, ointments or pasts; or solutions or suspensions such as drops. Formulations for topical administration to the skin surface can be prepared by dispersing the drug with a dermatologically acceptable carrier such as a lotion, cream, ointment or soap. Particularly useful are carriers capable of forming a film or layer over the skin to localize application and inhibit removal. For topical administration to internal tissue surfaces, the agent can be dispersed in a liquid tissue adhesive or other substance known to enhance adsorption to a tissue surface. For example, hydroxypropylcellulose or fibrinogen/thrombin solutions can be used to advantage. Alternatively, tissue-coating solutions, such as pectin-containing formulations can be used.

For inhalation treatments, inhalation of powder (self-propelling or spray formulations) dispensed with a spray can, a nebulizer, or an atomizer can be used. Such formulations can be in the form of a fine powder for pulmonary administration from a powder inhalation device or self-propelling powder-dispensing formulations. In the case of self-propelling solution and spray formulations, the effect may be achieved either by choice of a valve having the desired spray characteristics (*i.e.,* being capable of producing a spray having the desired particle size) or by incorporating the active ingredient as a suspended powder in controlled particle size. For administration by inhalation, the compounds also can be delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.,* a gas such as carbon dioxide, or a nebulizer.

Systemic administration also can be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants generally are known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and filsidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or

suppositories. For transdermal administration, the active compounds typically are formulated into ointments, salves, gels, or creams as generally known in the art.

The active compounds may be prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials also can be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811. Microsomes and microparticles also can be used.

Oral or parenteral compositions can be formulated in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

As noted above, drugs identified or designed according to the invention can be formulated into pharmaceutical compositions by admixture with pharmaceutically acceptable nontoxic excipients and carriers. Such compositions can be prepared for parenteral administration, particularly in the form of liquid solutions or suspensions; for oral administration, particularly in the form of tablets or capsules; or intranasally, particularly in the form of powders, nasal drops or aerosols. Where adhesion to a tissue surface is desired the composition can include the drug dispersed in a fibrinogen-thrombin composition or other bioadhesive. The drug then can be painted, sprayed or otherwise applied to the desired tissue surface. Alternatively, the drugs can be formulated for parenteral or oral administration to humans or other mammals, for example, in therapeutically effective amounts, *e.g.,* amounts which provide

appropriate concentrations of the drug to target tissue for a time sufficient to induce the desired effect.

Where the active compound is to be used as part of a transplant procedure, it can be provided to the living tissue or organ to be transplanted prior to removal of tissue or organ from the donor. The drug can be provided to the donor host. Alternatively or, in addition, once removed from the donor, the organ or living tissue can be placed in a preservation solution containing the active compound. In all cases, the active compound can be administered directly to the desired tissue, as by injection to the tissue, or it can be provided systemically, either by oral or parenteral administration, using any of the methods and formulations described herein and/or known in the art.

Where the drug comprises part of a tissue or organ preservation solution, any commercially available preservation solution can be used to advantage. For example, useful solutions known in the art include Collins solution, Wisconsin solution, Belzer solution, Eurocollins solution and lactated Ringer's solution.

The effective concentration of the compounds to be delivered in a therapeutic composition will vary depending upon a number of factors, including the final desired dosage of the compound to be administered and the route of administration. The preferred dosage to be administered also is likely to depend on such variables as the type and extent of disease or indication to be treated, the overall health status of the particular patient, the relative biological efficacy of the compound delivered, the formulation of the drug, the presence and types of excipients in the formulation, and the route of administration. In general terms, the drugs of this invention can be provided to an individual using typical dose units deduced from the earlier-described mammalian studies using non-human primates and rodents.

When the active compounds are nucleic acid molecules, the nucleic acid may be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or by stereotactic injection (see *e.g.,* Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from

recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

When an active compound of the invention is intended for administration to a plant host, the invention may be applied directly to the plant environment, for example, to the surface of leaves, buds, roots or floral parts. Alternatively, the present invention can be used as a seed coating. The determination of an effective amount of the present invention as required for a particular plant is within the skill of the art and will depend on such factors as the plant species, method of planting, and soil type. It is contemplated that compositions containing drugs of the invention can be prepared by formulating such drugs with adjuvants, diluents, carriers, etc., to provide compositions in the form of filings/divided particulate solids, granules, pellets, wetable powders, dust, aqueous suspensions or dispersions, and emulsions. It is further contemplated to use such drugs in capsulated form, for example, the drugs can be encapsulated within polymer, gelatin, lipids or other formulation aids such as emulsifiers, surfactants wetting agents, antifoam agents and anti-freeze agents, may be incorporated into such compositions especially if such compositions will be stored for any period of time prior to use. Application of compositions containing drugs of the invention as the active agent can be carried out by conventional techniques. When an active compound is intended for administration to an insect host, standard methods such as, but not limited to, aerial dispersal are contemplated.

Active compound identified or designed by a method of the invention also include precursors of the active compounds. The term precursors refers to a pharmacologically inactive (or partially inactive) derivative of a parent molecule that requires biotransformation, either spontaneous or enzymatic, within the organism to release the active compounds. Precursors are variations or derivatives of the compounds of the invention which have groups cleavable under metabolic conditions. Precursors become the active compounds of the invention which are pharmaceutically active in vivo, when they undergo solvolysis under physiological conditions or undergo enzymatic degradation. Precursor forms often offer advantages of solubility, tissue compatibility, or delayed release in the mammalian organism (see Bundgard, Design of Prodrugs, pp. 7-9, 21-24, Elsevier, Amsterdam (1985); and Silverman, The Organic Chemistry of Drug Design and Drug Action, pp. 352-401, Academic Press, San Diego, CA (1992).

Active compound as identified or designed by the methods described herein can be administered to individuals to treat disorders (prophylactically or therapeutically). In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a drug as well as tailoring the dosage and/or therapeutic regimen of treatment with the drug.

With regard to mammals, it is contemplated that the effective dose of a protein synthesis inducer or inhibitor will be in the range of about 0.01 to about 50 mg/kg, preferably about 0.1 to about 10 mg/kg of body weight, administered in single or multiple doses. Typically, the inducer or inhibitor may be administered to a human recipient in need of treatment at a daily dose range of about 1 to about 2000 mg per patient.

In light of the foregoing general discussion, the specific examples presented below are illustrative only and are not intended to limit the scope of the invention. Other generic and specific configurations will be apparent to those persons skilled in the art.

III. Examples

A. Example 1: Preparation of 50S Ribosomal Subunit Crystals

H. marismortui (ATCC 43049) was grown as described previously (Ban *et al.* (1998) *supra*) on a slightly modified version of ATCC culture medium 1230, which was supplemented with 4.3 g of yeast extract, 5.1 g of Tris, and 3.4 g of glucose per liter. Bacteria were grown at 37°C to an OD_{550nm} between 1.0 and 2.2. They were harvested by centrifugation, and stored at -80°C. Cells were ruptured using a French press. Ribosomes were prepared from lysates by centrifugation, and subunits were isolated on sucrose gradients (Shevack *et al.* (1985) *FEBS Lett.* 184: 68-71).

1. Reverse Extraction

(1) Take 1 mg of subunits from a concentrated 50S ribosomal subunit stock (30mg/ml in 1.2 M KCl, 0.5 M NH4Cl, 20 mM $MgCl_2$, Tris 10 mM, $CdCl_2$ 1 mM, Tris 5 mM, pH 7.5) and mix with 1/2 vol. of 30% PEG6000 (300g PEG, 700 ml H_20 to make 1 liter of 30% PEG; filter through 0.2 μm filter). Leave on ice for 1 to 2 hr.

(2) Spin down precipitate for about 30 seconds using a desktop centrifuge.

(3) Remove supernatant and add 100 μl of RE-buffer: 7% PEG6000, 1.2 M KCl, 0.5 M NH_4Cl, 100 mM KAc, 30 mM $MgCl_2$, 10 mM Tris, 10 mM MES (pH 7.5), and 1 mM $CdCl_2$

(4) Resuspend pellet at room temperature by mixing with a P200 pipette set at 50 μl. Resuspended material should appear a little cloudy.

(5) Wrap the eppendorf tube in aluminum foil and leave for equilibration at room temperature for 30-60 min. The solution will be saturated with 50S.

(6) Spin down for 2 minutes in desk-top centrifuge at room temperature, transfer supernatant to new eppendorf tube. A little pellet should be found in the tube used for centrifugation. Keep the supernatant at room temperature.

(7) Put 8-10 μl of supernatant in the sample well of a sitting drop tray (Charles-Supper). Streak seed one hour later from a seed stock. Seed stock is prepared by putting previously grown crystals in stabilizing solution buffer A (see below), and then vortexing them violently. To streak seed, a human hair cleaned with water and ethanol and then dried is passed through the vortexed solution and then touched on the new crystallization drop. Drops should look cloudy. The reservoirs in the sitting drop trays contain 1000 μl of a solution containing 8% PEG6000, 1.2 M KCl, 0.5 M NH_4Cl, 100 mM KAc, 6.5 mM HAc (yields pH 5.8), 30 mM $MgCl_2$, and 1 mM $CdCl_2$.

(8) Check after one day if seeding is succeeded. If yes, let crystals grow for three weeks.

2. Stabilization Protocol

When crystals have finished growing (after approximately 3 weeks), each sitting drop chamber is opened by making just a single cut (slit) going from the middle and to the edge of the well. Through this narrow slit, 10 µl of buffer A (1.2 M KCl, 0.5 M NH_4Cl, 30 mM $MgCl_2$, 10% PEG6000, 1 mM $CdCl_2$, 100 mM KAc, 10 mM Tris (titrated to final pH 6.1), 30 mM MES) at room temperature is added to each drop and 45 µl of Buffer C (0.667 M MES, 0.333 M Tris) to each reservoir.

Trays are put in a plastic box with a lid, and put in a 16° C incubator for approximately one day, and then moved to 12° C for another day. The plastic box is then put in a polystyrene container with a lid, and put in the cold room for yet another day. Crystals can be kept like this for a long time, but need to undergo a further changing of buffer prior to any use.

Make the following transition series using buffer A and buffer B (1.7 M NaCl, 0.5 M NH_4Cl, 30 mM $MgCl_2$, 1 mM $CdCl_2$, 12% PEG6000, 20% EG, 100MM KAC (titrated to final pH 5.8 with HAC) to give final ratios of buffer B to buffer A of: 1/16, 1/8, 1/4, 1/2, 3/4. All solutions should be at cold room temperature. All manipulations of the drops will take place through the narrow slit.

(1) Add 40 µl "1/16" to the drop, leave for 15 minutes.

(2) Add 40 µl "1/8" to the drop, leave for 30-60 minutes.

(3) Take out 40 µl from the drop (and discard it in the reservoir), add 40 µl "1/4", leave for 30-60 minutes.

(4) Take out 40 µl from the drop (and discard it in the reservoir), add 40 µl "1/2", leave for 15 minutes.

(5) Take out 40 µl from the drop (and discard it in the reservoir), add 40 µl "3/4", leave for 15 minutes.

(6) Take out 40 µl from the drop (and discard it in the reservoir), add 40 µl buffer B, leave for 15 minutes.

(7) Take out 60-80 µl from the drop (and discard it in the reservoir), add 60-80 µl Buffer B, replace reservoirs with 500 µl buffer B.

B. Example 2: Determination of the Crystal Structure of the 50S Ribosomal Subunit, With the Initial Refinement

All data, except the two native data sets, were collected at the National Synchrotron Light Source (Brookhaven) from crystals frozen at 100 K, using beamlines X12b and X25 and recorded using a 345 mm MAR imaging plate. For each heavy atom derivative, anomalous diffraction data were collected at the wavelength corresponding to the peak anomalous scattering. The beam size was 100 x 100 µm for most data collections at X25 and 200 x 200 µm at beamline X12b. The crystals were aligned along the long axis of the unit cell (570 Å) so that 1.0° oscillations could be used to collect reflections out to a maximum of 2.7 Å resolution at the edge of the MAR detector. At beamline X12b the crystal to detector distances varied between 450.0 mm and 550.0 mm depending on wavelength, crystal quality and beam divergence, and it was chosen so that maximum resolution data could be collected while avoiding overlapping of spots. At beamline X25 the detector was positioned on a rigid platform at 480 mm which allowed data collection to 3.2 Å for iridium and osmium derivatives with the wavelength set at the anomalous edge. Native data to 2.4 Å resolution were collected at the structural biology beamline ID19 of the Advanced Photon Source (Argonne) using a CCD detector (*etc.*). Data sets were processed by using DENZO and SCALEPACK (Otwinowski, (1993) Data Collection and Processing).

Heavy atom based phasing was extended to 3.2 Å resolution by combining MIR phases calculated for two different isomorphous groups of data (MIR1 and MIR2, Table 1) with single derivative anomalous dispersion (SAD) phases. The best two derivatives were osmium pentamine and iridium hexamine, each of which contained a large number of binding sites (Table 1). Several other derivatives with smaller number of sites further improved map quality. All phasing was done by maximum likelihood method implemented in CNS (Brünger *et al.* (1998) *supra*) with the exception of the Ta_6Br_{12} derivative, which was refined in SHARP (de La Fortelle, (1997) *Meth. Enzymol.* 276: 472-494) represented as spherically averaged electron density (Table 1). Phases were improved and extended from 3.3 Å to 2.4 Å by solvent flipping (Abrahams *et al.* (1996) *supra*) and models were built.

C. Example 3: Preparation of Crystals of 50S Ribosomal Subunit/Puromycin Complex and Collection of X-ray Diffraction Data

Crystals of 50S ribosomal subunits were grown and stabilized as described earlier. CCdA-p-puromycin (see Figure 9A) was a generous gift from Michael Yarus (Welch, *et al.* (1995) *supra*). Oligonucleotides from amino-N-acylated minihelices (see Figure 9B) were synthesized by Dharmacon. Following deprotection, the oligonucleotides were heated briefly to 100° C and snap-cooled on ice to reanneal. Ribosomal 50S subunit crystals were stabilized and then soaked for 24 hours in stabilization buffer plus 100 μM CCdA-p-puromycin or amino-N-acylated mini-helices prior to cryovitrification in liquid propane and X-ray diffraction data collection. Phases were calculated by density modification (CNS) beginning with the best experimental phases using $2F_o$(analogue)- F_o(native) for amplitudes, from 60.0 to 3.2 Å. (Native amplitudes were from the most isomorphous native 1 data set, except for those amplitudes which were present only in the more complete native 2 data set. Calculated $2F_o$-F_o amplitudes which were less than twice the corresponding calculated & were replaced by F_o(analogue)). Maps were then calculated using phases from density modified and $2F_o$(analogue)- F_o(native) or F_o(analogue)- F_o(native) amplitudes.

D. Example 4: Antibiotic Binding Sites Located in the Polypeptide Exit Tunnel Near the Peptidyl Transferase Center

Crystalline complexes of the *H. marismortui* large subunit complexed with three antibiotics have been established at about 3.0 Å resolution. The electron density maps at this resolution have allowed us to position approximately on the ribosome the antibiotics tylosin, carbomycin and anisomycin. We observed that these antibiotics all bind to the ribosome in the region that lies between the peptidyl transferase center as defined by the Yarus inhibitor, CCdA-p-puromycin, and the tips of the proteins L22 and L4 at the point that they form a small orifice in the polypeptide exit tunnel. The general location of this major antibiotic binding site is shown in Figure 19. Tylosin and carbomycin appear to function by blocking the exit of newly synthesized polypeptides. Anisomycin blocks the A site.

It is contemplated that the antibiotic erythromycin will bind in almost the same location as tylosin because of the similarity of the two molecules and because erythromycin resistance

mutations are known in both the tip of protein L4 and in portions of the RNA near the tylosin binding site.

The vast majority of the interactions between those antibiotics and the ribosome are through rRNA that forms the A site, and the surface of the tunnel between the peptidyl transferase center and protein L22. Since these antibiotics do not bind identically, there will be many additional ways that small molecule compounds can be designed to bind in this region using the tools and methodologies of the present invention. For example, by connecting together components of each of the different antibiotics which bind to non-overlapping sites it will be possible to create new antibiotics (see, Example 6). In addition, based on new principles of small molecule RNA interaction shown by these antibiotic complexes we will be able to design entirely novel small molecules that will bind to sites on the ribosome as well as other potential RNA targets.

E. Example 5: Design and Testing of Hybrid Antibiotics

Many antibiotics that target ribosomes, more particularly large ribosomal subunits, and disrupt protein synthesis are complex molecules that are effectively concatenations of simpler substructures, at least one of which interacts with a discrete part of the ribosome. When the compound in question includes several interactive substructures, its binding site is effectively the sum of the subsites that contact and engage each such substructure. It has been found that many antibiotics that target the large ribosomal subunit bind the ribosomal subunit at sites that are close to one another. Thus the possibility exists of synthesizing new antibiotics in which one ribosome-binding moiety of a first known antibiotic is linked chemically to a ribosome-binding moiety of a second known antibiotic that interacts with an adjacent subsite. The new compound that results is thus a chimera of the two antibiotics from which it derives.

Chimeric antibiotics can be designed using the information about the structures of antibiotic/ribosome complexes discussed hereinabove. These structures permit the identification of antibiotic binding subsites in the ribosome, and the specification of the chemical entities that interact with them. Equipped with such knowledge, those skilled in the art of organic synthesis can synthesize compounds that link the substructures of interest together in ways that should enable them to interact with their respective subsites at the same time. Any compound devised

this way that functions in the manner intended is likely to inhibit cell growth and if it does, protein synthesis *in vivo*. At the very least, it should block protein synthesis in *in vitro* assay systems. Further information about the ribosomal interactions of such a compound can be obtained by determining the structure of the complex it forms with the ribosome using the methods described in Section D, hereinabove.

For example, as a result of the work described herein, it has been discovered that the disaccharide moiety of carbomycin binds the large ribosomal subunit at a site in close proximity to the binding site for a portion of the anisomycin. Using this information and the software packages described hereinabove, the skilled artisan can design a hybrid antibiotic comprising the relevant ribosome binding portions of carbomycin and anisomycin linked by a suitable chemical linker. This hybrid molecule, once designed, can be synthesized and purified using conventional synthetic organic chemistries and conventional purification schemes. Once synthesized and purified, the hybrid molecule can be screened for bioactivity. These screens can include, for example, growing micro-organisms on or in media either supplemented or lacking the hybrid molecule. Any reduction in the number of micro-organisms or the size of colonies in the presence of the hybrid molecule would be indicative of bioactivity. Furthermore, the hybrid molecule could be tested in a cell free translation system in the presence of one or more labeled amino acids. Any reduction in the level of labeled amino acids incorporated into proteins in cell free systems that include the hybrid molecule relative to cell free systems locking the hybrid molecule would be indicative that the hybrid molecule acts as a functional protein synthesis inhibitor. It is contemplated that the hybrid molecule could then be iteratively refined as discussed hereinabove to enhance its bioactive peptides and bioavailability.

INCORPORATION BY REFERENCE

The disclosure of each of the patent documents, scientific articles, atomic-co-ordinates (including, without limitation, those sets deposited at the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) with the accession numbers PDB ID: 1FFK; PDB ID: 1FF2; PDB ID: 1FG0; and PDB ID: 1JJ2, and/or contained on Disk No. 1, 2, or 3) referred to herein is incorporated by reference herein.

All materials submitted herewith on Disk Nos. 1, 2 and 3 are incorporated by reference herein. Disk Nos. 1, 2 and 3 are identified as containing the following files:

Disk No. 1 of 3:

File Name:	Size:	Date Created
1) 1ffk.doc	7,046KB	7/25/01
2) 1ffk.ent	5,484KB	7/25/01
3) 1ffz.doc	1,219KB	7/25/01
4) 1ffz.ent	937KB	7/25/01
5) 1fg0.doc	1,225KB	7/25/01
6) 1fg0.ent	942KB	7/25/01
7) <dir>		7/25/01
8) <dir>		7/25/01

Disk No. 2 of 2:

File Name:	Size:	Date Created:
1) 1jj2.rtf	12,742KB	7/25/01
2) 1jj2.txt	8,372KB	7/26/01
3) <dir>		7/26/01
4) <dir>		7/26/01

Disk No. 3 of 3:

File Name:	Size:	Date Created:
1) anisomycin.pdb	7,593KB	7/18/01
2) blasticidin.pdb	7,594KB	7/18/01
3) carbomycin.pdb	7,592KB	7/18/01
4) sparsomycin.pdb	7,541KB	7/18/01
5) spiramycin.pdb	7,592KB	7/18/01
6) tylosin.pdb	7,592KB	7/18/01
7) virginiamycin.pdb	7,591KB	7/18/01
8) <dir>		7/18/01
9) <dir>		7/18/01

EQUIVALENTS

The invention may be embodied in other specific forms without departing form the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.